Health Insurance and the Demand for Medical Care: Instrumental Variable Estimates using Health Insurer Claims Data*

Abe Dunn†

May 24, 2013

Abstract

Understanding the effect of price on medical care demand is a critical economic question, as policy-makers search for ways to stem the rapid growth in medical expenditures. However, estimating demand is particularly challenging because individuals self-select into insurance plans. This paper takes a different approach to estimating demand that uses the negotiated prices between insurers and providers as an instrument. The instrument is viewed as a textbook "cost shifting" instrument that impacts plan offerings, but is unobserved by consumers. The paper finds a price elasticity of demand of around -0.20, matching the elasticity found in the RAND health insurance experiment.

1 Introduction

U.S. medical care expenditures account for a large and growing share of GDP and policy-makers continue to search for mechanisms to rein in expenditure growth. In this environment, understanding the demand for medical care is critical. Estimates of the price elasticity of demand may improve our understanding of patient incentives and lead to policies to help slow the growth of the health care sector. Unfortunately, estimating medical care demand is particularly challenging. One of the central problems is that the marginal price of medical care faced by consumers is often determined by consumers through their selection of a health insurance plan. For instance, the least healthy individuals may be more likely to choose a plan with the most generous insurance coverage, leading to an overestimate of medical care demand elasticity when looking at correlations between the out-of-pocket price and the utilization of medical care.

Both the economic importance of measuring the elasticity of demand as well as the substantial empirical challenge caused by selection were key motivations for conducting the RAND health insurance experiment in the 1970s. The RAND experiment was specifically designed to address the selection problem. The key to its success was the randomization of health insurance coverage across the sample population that allowed researchers to side-step the selection issue and isolate the effect of cost sharing on demand. Although it has been more than 30 years since the RAND experiment was conducted, it remains the gold standard for

*The views expressed in this paper are solely those of the author and do not necessarily reflect the views of the Bureau of Economic Analysis.

†Bureau of Economic Analysis; abe.dunn@bea.gov

understanding consumer responsiveness to out-of-pocket price. However, the study has several limitations. Most importantly, since the study was conducted, the share of GDP devoted to medical care has doubled and medical technologies have changed substantially. These dramatic changes suggest that the evidence from the RAND experiment may be relatively dated and there are also questions regarding medical care demand that remain unanswered in today's environment.[1] Consequently, researchers have continued to search for alternative approaches to estimating the demand for medical care.

This paper takes a different approach to estimating demand, which relies on an often noted industry feature: the out-of-pocket price paid by the consumer is typically not the same as the full price paid to the medical care provider (i.e., the allowed amount). With this in mind, this paper argues that the negotiated price between insurers and medical providers in an MSA may be thought of as a textbook "cost shifter" instrument. The theoretical justification is clear: the package of benefits offered to enrollees will be affected by profit maximizing insurers responding to the negotiated price for medical services in an area. At the same time, the negotiated price should be uncorrelated with the selection of an insurance plan, since consumers are typically unaware of the negotiated prices with providers.[2] Moreover, medical provider contracts are negotiated prior to setting insurance plan offerings and the negotiated price is typically the same for both the least generous plans and the most generous plans, greatly reducing the possibility that the instrument would be related to plan selection. Finally, the instrument is likely to be strong, since the negotiated price differs substantially across MSAs. This empirical fact is documented in detail by Dunn, Shapiro, and Liebman (2012). This can also be seen by looking at examples of specific price differences. For instance, the average negotiated price for a 15-minute office visit with a general MD in Minneapolis, MN, in 2007 is \$82, while in Memphis, TN, the average is \$63.[3]

This instrumental variable (IV) strategy is fundamentally different from prior work. To control for endogeneity, researchers typically look for factors that affect the out-of-pocket price that are unrelated to the demand for insurance. This may be caused by randomness from an actual experiment,[4] a natural experiment,[5] or through another instrument that is related to the marginal price faced by a consumer, but unrelated to insurance selection.[6] In contrast, the identification strategy in this paper focuses on how changes in the underlying marginal cost of medical services affect the incentives of insurers, which ultimately impacts the out-of-pocket prices faced by consumers. While this approach is unique to the estimation of medical care demand, this basic intuition is often the motivation behind instrumental variable strategies applied in the industrial organization literature (e.g., Hausman (1996) and Nevo (2001)).

The demand model is estimated using individual micro data from the MarketScan commercial claims database for the years 2006 and 2007. The MarketScan data is a convenience sample of enrollees from insurers and large employers. The data includes the demographic information of individuals, such as the age, sex, and type of insurance plan. Most importantly, the data includes information on the medical

[1] Addressing these issues by conducting another experiment may be very costly. Manning et al. (1987) report costs of a little more than \$136 million in 1984 dollars or \$408 million in inflation-adjusted 2013 dollars. Even if another experiment is conducted, unique empirical challenges also arise in an experimental setting (see Aron-Dine, Einav, and Finkelstein (2012)).

[2] This fact was highlighted in great detail in the *Time* magazine article "Bitter Pill: Why Medical Bills Are Killing Us" by Steven Brill.

[3] These estimates were computed using MarketScan data described later in the paper. Similar differences are also found looking at median price differences.

[4] e.g., the RAND study (see Manning et al. (1987) and Keeler and Rolph (1988)).

[5] e.g., see Phelps and Newhouse (1972), Cherkin, Grothaus and Wagner (1989), and Selby, Fireman and Swain (1996). More recently, the Oregon Health Insurance Experiment (see Finkelstein et al. (2012) and Baicker et al. (2013)).

[6] e.g., Kowalski (2010) and Duarte (2012).

conditions of the enrollees, utilization of medical care services, and expenditures. The expenditure data indicates both the amount paid out-of-pocket by the enrollee and the total allowed amount paid to the providers. Data on income, education, and health are also incorporated into the analysis.

In addition to the basic features of the data just mentioned, the MarketScan data is extremely detailed and large, with more than four million enrollees in each year. These unique aspects of the data are essential for constructing an instrument that accurately reflects the marginal cost of insurers. The instrument is computed by building an index that isolates the variation in underlying service prices (for example, the negotiated price for of a MRI for a patient with back pain), but holding utilization constant (for example, fixing the number of MRIs for treating back pain). Accurately constructing a service price index across many MSAs requires a significant amount of detailed information, since physicians and hospitals offer an enormous number of products and services.

The main result of the paper is that the individual price elasticity of medical care utilization is about -0.20, which matches the estimate found in the RAND study. Following the RAND study, this paper looks at price responsiveness at the disease episode level, investigating the effect of price on the intensive margin (i.e., utilization per disease episode) and the extensive margin (i.e., the number of episodes). Similar to the RAND study, price responsiveness on the intensive margin accounts for only a small fraction of the total elasticity. Most of the individual responsiveness to the out-of-pocket price is on the number of episode occurrences. These findings confirm the relevance of the RAND estimates in the current environment and outside of the experimental setting. Overall, the methodology and empirical findings in this paper are of general interest as they uncover a new way of identifying consumer responsiveness from real world price movements.

Although this paper argues that the negotiated service price in the MSA is a valid instrument, much of the analysis focuses on the potential for endogeneity to creep into the negotiated price in an MSA. For example, a bias could potentially enter the model if the service price in an MSA is related to the quality of services in the MSA. For this reason, a variety of IV strategies are employed, although the basic idea behind each strategy is the same – to find a variable related to the marginal costs of insurer generosity. Following arguments similar to Hausman (1996), one alternative IV strategy uses the service price indexes from other MSAs within the same state. As another IV strategy, the demand for medical care services for those individuals enrolled in one plan type (e.g., PPO plans) are instrumented by using the negotiated service prices for individuals enrolled in another plan type (e.g., POS plans). While one may think that these patterns may be related to cross-market provider practices, controlling for per capita Medicare expenditures has no effect on these findings. Other IV strategies and many robustness checks are analyzed and under several alternative specifications the main results of the paper remain qualitatively unchanged.

The next section discusses the construction of the price and utilization measures. Section 3 describes the empirical model. Section 4 presents the data and descriptive statistics. Section 5 presents results and section 6 concludes.

2 Defining Service Prices and Utilization

The analysis in this section relies on many of the basic ideas presented in Dunn, Shapiro and Liebman (2012). To begin thinking about measuring medical care utilization and prices, it is helpful to start with a simple example. Suppose there is just a single patient, i, that is seeking treatment for high blood pressure, often referred to as hypertension (h). For simplicity, the example will start by supposing that there is only one type of treatment available, the treatments are 15-minute office visits where the patient's blood pressure is monitored.[7] Let

$$\begin{aligned} c_{h,i} &= \text{All expenditures incurred for high blood pressure} \\ & \text{(i.e., out-of-pocket expenditures plus expenditures paid by the insurer).} \\ q_{h,i} &= \text{Number of 15-minute visits with the physician.} \\ p_{h,i} &= \text{Price per 15-minute visit with the physician (i.e., } \frac{c_{h,i}}{q_{h,i}}\text{).} \end{aligned}$$

Also suppose that there is a reference or base group, B, so that $c_{h,B}$, $q_{h,B}$, and $p_{h,B}$ are the total expenditures, number of 15-minute visits, and price for 15-minute visits for this base group. In this example the individual service price ($SP_{h,i}$) for person i may be calculated as: $SP_{h,i} = \frac{p_{h,i} \cdot q_{h,B}}{p_{h,B} \cdot q_{h,B}} = \frac{p_{h,i}}{p_{h,B}}$. This measures the contracted price per 15-minute visit relative to the base group's price. Differences in $SP_{h,i}$s across patients would reflect only differences in the contracted prices, not the number of visits. Dividing this $SP_{h,i}$ into the total expenditure of the episode ($c_{h,i}$) gives the utilization measure. That is, the individual service utilization is $SU_{h,i} = \frac{c_{h,i}}{SP_{h,i}} = p_{h,B} \cdot q_{h,i}$. This utilization measure indicates how much the insurer and patient would have paid in total for the patient's, $q_{h,i}$, 15-minute visits if the contracted price were equal to the base group price. Differences in $SU_{h,i}$s across patients reflect only differences in the number of 15-minute visits. To think about this utilization measure in terms of indexes, the total expenditures for patient i relative to the base group may be written as the product of a price index and a utilization index.

$$\frac{c_{h,i}}{c_{h,B}} = \left(\frac{p_{h,i} \cdot q_{h,i}}{p_{h,B} \cdot q_{h,i}}\right) \cdot \left(\frac{p_{h,B} \cdot q_{h,i}}{p_{h,B} \cdot q_{h,B}}\right) \tag{1}$$

The first term in equation (1) is a price index, and the second term is a utilization index. Ignoring the fixed denominator in the utilization index ($p_{h,B} \cdot q_{h,B}$), the numerator is the individual service utilization measure, $SU_{h,i}$. While this example focuses on one precisely defined procedure, clearly physicians perform many alternative types of procedures other than 15-minute office visits. More generally, let $q_{h,i}$ be a measure of the amount of services performed, where the total amount paid is calculated by multiplying the service price times utilization, $p_{h,i} \cdot q_{h,i}$. The precise calculation of the amount of services, $q_{h,i}$, will be discussed in greater detail in the data section of the paper. For those familiar with medical care payments, this measure of utilization may be thought of as a relative value unit, which reflects the amount of services performed and is typically used when calculating payments to physicians.

Expanding on this example, now suppose that this hypertension patient may be treated with two types of services, prescription drug and physician office services, where the service categories correspond to the subscripts (D) and (O). That is, $q_{h,i,O}$ and $p_{h,i,O}$ are the utilization and price for the physician office visits, and $q_{h,i,D}$ and $p_{h,i,D}$ are the utilization and price for prescription drugs. Continuing with the index decomposition that is parallel to (1), but with two services, the decomposition becomes:

[7] This type of procedure may fall under the specific service code 99213 as defined by the Current Procedure Terminology (CPT) code.

$$
\begin{aligned}
\frac{c_{h,i}}{c_{h,R}} &= \frac{p_{h,i,O} \cdot q_{h,i,O} + p_{h,i,D} \cdot q_{h,i,D}}{p_{h,B,O} \cdot q_{h,B,O} + p_{h,B,D} \cdot q_{h,B,D}} \\
&= \left(\frac{p_{h,i,O} \cdot q_{h,i,O} + p_{h,i,D} \cdot q_{h,i,D}}{p_{h,B,O} \cdot q_{h,i,O} + p_{h,B,D} \cdot q_{h,i,D}} \right) \cdot \left(\frac{p_{h,B,O} \cdot q_{h,i,O} + p_{h,B,D} \cdot q_{h,i,D}}{p_{h,B,O} \cdot q_{h,B,O} + p_{h,B,D} \cdot q_{h,B,D}} \right)
\end{aligned} \tag{2}
$$

The second term of the decomposition is a utilization index, and the numerator of the index corresponds to the service utilization variable studied in this paper: $SU_{h,i} = p_{h,B,O} \cdot q_{h,i,O} + p_{h,B,D} \cdot q_{h,i,D}$.

The general case follows from this basic example. The medical care expenditure for the treatment of a disease episode is defined as the total dollar amount of medical care used until treatment is completed, including *all* service categories.[8] Formally, denote the expenditure paid to medical providers for an episode of treating disease d for insurance enrollee i as $c_{d,i}$. The individual disease expenditure, $c_{d,i}$, can be divided between service price and service utilization components. This can be seen by showing that the expenditure is calculated by totaling dollars spent on all services: $c_{d,i} = \sum_s p_{d,s,i} q_{d,s,i}$ where $q_{d,i,s}$ and $p_{d,i,s}$ are the service utilization and service price components for diseases episode d for individual i for service type s. Following the examples, to obtain an individual service utilization measure, the base service price for service type s, $p_{d,B,s}$, is multiplied by utilization amounts for different services:

$$SU_{d,i} = \sum_s q_{d,i,s} \cdot p_{d,B,s}. \tag{3}$$

An individual may have more than one disease episode. For instance, an individual may have diabetes, hypertension, and heart disease. An overall utilization measure may be calculated by summing the disease-specific utilization measure over the different disease episodes for individual i:

$$SU_i = \sum_{d \in i} SU_{d,i}. \tag{4}$$

One can divide this measure of overall utilization into two distinct pieces: the amount of utilization per episode (i.e., the intensive margin) and number of disease episodes (i.e., the extensive margin). The conceptual justification for measuring utilization along two dimensions is that the physician's influence along the intensive margin and extensive margin may be quite distinct. The patients may choose to seek care with a physician to treat their medical conditions, but after seeking treatment, the patient may have less control over the intensity of treatment recommended by the physician.

While $SU_{d,i}$ is the measure of utilization per episode, the number of episodes can be calculated by summing the number of disease episodes for each enrollee i (i.e., $Episodes_i = \sum_{d \in i} 1$).[9] However, this simple count may not accurately reflect the large differences in the intensity of treatment across disease episodes. For example, the average intensity of treatment for hypertension is much lower than that of ischemic heart disease. Specifically, let the average utilization measure for disease d be calculated as, $\overline{SU}_d = \frac{\sum_i SU_{d,i}}{\text{Number of individuals with disease } d}$. Then it should be expected that $\overline{SU}_{\text{heart disease}} > \overline{SU}_{\text{hypertension}}$. To construct a disease episode count that reflects the different average intensities across disease episodes, a

[8] For example, for an individual with a broken foot, the episode of treatment will be defined by the dollar of medical services used to treat that condition from the first visit to a provider until the foot is healed. For medical conditions that are chronic, we interpret an episode as expenditure for services used to treat the chronic condition over a one year period.

[9] If an enrollee has multiple disease episodes of the same type, this will be counted as multiple episodes. For instance, an individual may have two episodes of a sore throat.

measure of the weighted number of episodes is calculated by summing over the average utilization amounts for each disease d of individual i,

$$Episodes_i^W = \sum_{d \in i} \overline{SU}_d. \tag{5}$$

The weighted number of episodes will provide the main unit of analysis for studying demand along the extensive margin. Note that the weighted number of episodes is unresponsive to changes in the amount of utilization per episode. For instance, if an individual has hypertension treated more intensively than average, this will have no effect on $Episodes_i^W$. The only factors that affect $Episodes_i^W$ are the number of disease episodes and the average intensity of those episodes, as measured by $\overline{SU}_d$.

The key explanatory variable in this study is the out-of-pocket price. Let $oope_{d,i}$ be the total out-of-pocket expenditures for individual i for disease episode d. The out-of-pocket price is just the out-of-pocket expenditure divided by utilization. Specifically, the equation used to compute an individual's out-of-pocket price ($OOPP$) is

$$OOPP_i = \frac{\sum_{d \in i} oope_{d,i}}{SU_i}. \tag{6}$$

For individuals enrolled in family plans, the average out-of-pocket price across all individuals i in family f is $OOPP_f = \frac{\sum_{i \in f} \sum_{d \in i} oope_{d,i}}{\sum_{i \in f} SU_i}$. The main analysis will focus on the average out-of-pocket price faced by the family, $OOPP_f$.[10] Some of the analysis in the following sections involves the calculation of individual-specific service price indexes that are constructed in a manner similar to $OOPP_i$. In particular, the individual service price (SP_i) may be calculated by summing over all individual expenditures (rather than the out-of-pocket expenditures) and dividing by the overall utilization measure: $SP_i = \frac{\sum_{d \in i} c_{d,i}}{SU_i}$.[11]

A nice feature of the out-of-pocket price measure is that identical services are priced similarly across markets. For example, if the out-of-pocket expenditure for a 15-minute office visit in city A is \$10 and the out-of-pocket expenditure for an identical 15-minute office visit in city B is \$15, then the out-of-pocket price measure in this paper would imply that the price for city B is 50 percent larger than city A (\$15/\$10=1.5) because the amount of utilization is the same, but the expenditure is 50 percent larger. In contrast, using a cost-sharing measure as the relevant price would not necessarily satisfy this property. For example, if the service price in city A were \$50 and the service price in city B were \$75, then the out-of-pocket prices implied by a cost-sharing measure in the two cities would be identical (i.e., $\frac{\$10}{\$50} = \frac{\$15}{\$75}$). Therefore, an attractive property of the out-of-pocket price measure, $OOPP_f$, is that the price is measured relative to a precisely defined unit of utilization, so that two different payment amounts for the same service will imply different price levels. As can be seen by this example, a very detailed data set is necessary to accurately price specific services and products (e.g., the methodology will need to distinguish between a 15-minute office visit, a 30-minute office visit, and an MRI).

2.1 MSA Service Price Index

An approach analogous to that described for measuring individual prices is taken to construct an MSA service price index. The average expenditure per episode of treating disease d in MSA r is denoted c_d^r. Similar to the individual level episode expenditures, the average expenditure, c_d^r, can be divided between service price and service utilization components. This can be seen more easily by showing that the average

[10] Alternative measures of out-of-pocket price are explored in robustness checks discussed later in the paper.

[11] Note that this corresponds to the price component of the index in (2)

expenditure per episode is calculated by totaling dollars spent on all services to treat the condition and dividing those dollars by the number of episodes: $c_d^r = \sum_s p_{d,s}^r Q_{d,s}^r / N_d^r$, where $Q_{d,s}^r$ is the quantity of services of type, s; $p_{d,s}^r$, is the service price; and N_d^r is the number of episodes treated.

To simplify notation, let q_d^r be a vector of the average amount of services utilized for the treatment of disease d in an MSA r, $q_d^r = Q_d^r / N_d^r$, where the component of the utilization vector for service s is , $Q_{d,s}^r / N_d^r$.[12] Also, let p_d^r be a vector of service prices, where the component of the vector for service s is, $p_{d,s}^r$. The price for a particular service type and disease can be calculated by dividing its average expenditure per episode for service s by the average utilization for service s: $p_{d,s}^r = \frac{c_{d,s}^r}{q_{d,s}^r}$ where $c_{d,s}^r$ is the average expenditure on disease d for service s in MSA r. For example, the price of an inpatient stay for treating heart disease is the total expenditure of an inpatient treatment for heart disease in an MSA, divided by the quantity of inpatient services for heart disease in that MSA.

This decomposition allows for an MSA service price index (SPI_d^r) for disease d in MSA r that is calculated as:

$$SPI_d^r = \frac{p_d^r \cdot q_d^B}{p_d^B \cdot q_d^B}, \tag{7}$$

which holds the utilization of services fixed at a base level.

This MSA service price index forms the basis for the main instruments used in this paper. The service price index is intended to capture the expected marginal cost for an additional unit of a medical care services for the typical enrollee in the population. Specifically, assuming full insurance, the SPI_d^r reflects the marginal cost of a service for treating a patient with disease, d, in MSA r relative to the base region, B. This service price index may also be viewed as the expected marginal cost of the next service. To see this, let the probability of receiving the next service from service type s be denoted $\Pr_{d,s}$, then the expected relative service price is $\sum_s \Pr_{d,s} \frac{p_{d,s}^r}{p_{d,s}^B}$. If the probability of each service is the expenditure share of the base group, $\Pr_{d,s} = \frac{p_{d,s}^B q_{d,s}^B}{p_d^B \cdot q_d^B}$, then the expected relative service price$=\sum_s \frac{p_{d,s}^B q_{d,s}^B}{p_d^B \cdot q_d^B} \frac{p_{d,s}^r}{p_{d,s}^B} = \frac{p_d^r \cdot q_d^B}{p_d^B \cdot q_d^B} = SPI_d^r$.

To calculate a service price index, SPI^r, that aggregates over diseases in MSA r, each disease-specific service price index, SPI_d^r, is weighted by the national expenditure share for that disease d for the entire U.S. Weighting by the expenditure share reflects the probability that the next dollar spent will be allocated to each disease.

3 Empirical Model of Demand

There are three distinct measures of utilization studied in this paper. First, the study focuses on the responsiveness to overall utilization, which looks at total medical care use, regardless of the disease being treated (i.e., SU_i). Second, similar to the RAND study, utilization is broken into two pieces: the number of episodes (i.e., $Episode_i^W$) and utilization per episode (i.e., $SU_{d,i}$). As argued by the RAND researchers (see Keeler and Rolph (1988)) and discussed briefly above, these two components of utilization likely involve different levels of control by physicians. The decision to treat an episode, such as hypertension or high cholesterol, may be thought of as a decision that is influenced by the consumer, while after initiating treatment, the physician may have relatively more control. In any case, for each of these measures of

[12] The services s are service categories, such as inpatient hospital or physician office services.

utilization the role of information and the relative control of the physician and the consumer will likely differ, which offers an important motivation for analyzing these decisions separately.

3.1 Components of Demand

3.1.1 Overall Utilization

To examine overall utilization, the overall utilization measure, SU_i, is regressed on the log of the out-of-pocket price, $\ln(OOPP_f)$, and individual demographics, Z_i. As is widely known in the health economics literature, medical care utilization may be highly skewed with a significant fraction of individuals with no utilization. To deal with these issues, this paper follows the guidelines outlined in the health econometrics literature to test functional forms and select the appropriate estimator. Following these guidelines, discussed in greater detail in the appendix, the main specification in this paper will apply a GLM model with a log link. Therefore, the empirical model of utilization is:

$$SU_i = \exp(\alpha \ln(OOPP_f) + \beta_1 Z_i + \delta \xi_i) + e_i,$$

where α and β_1 are parameters to be estimated and e_i is a random error term. The potential endogeneity of the out-of-pocket price variable is specified using the unobserved variable ξ_i. As an example, ξ_i may include unobserved illness severity, which may be related to both more generous insurance and the utilization of more services, creating a downward bias on α. In addition to an omitted variable problem, the out-of-pocket price may be measured with error. For example, the constructed out-of-pocket price measure, $OOPP_f$, may not match the marginal out-of-pocket price, as perceived by the consumer. Both the possibility of omitted variable bias and measurement error imply that it is important to apply an IV estimator.

The instrumental variable model applied in this paper is a two-stage residual inclusion model (a type of control function model).[13] The basic instrument used in this analysis is the MSA service price index, SPI^r. The first-stage regression of the IV procedure is:

$$\ln(OOPP_f) = \gamma \ln(SPI^r) + \tau_1 Z_i + u_i. \tag{8a}$$

To correct for endogeneity, the error term from the first-stage regression is included in the GLM model to control for the unknown factors causing movements in out-of-pocket prices, such as unobserved health and measurement error, and isolates those movements due to exogenous factors. Specifically, the estimate $\widehat{\xi}_i = \ln(OOPP_f) - (\widehat{\gamma} \ln(SPI^r) + \widehat{\tau}_1 Z_i)$ is included in the GLM model and the second-stage regression is

$$SU_i = \exp(\alpha \ln(OOPP_f) + \beta_1 Z_i + \delta \widehat{\xi}_i) + e_i. \tag{9}$$

3.1.2 Weighted Number of Episodes - Extensive Margin

The weighted number of episodes is studied in a similar fashion to overall utilization. The analysis changes by substituting the dependent variable SU_i in (9) with the weighted number of treated episodes, $Episodes_i^W$. A two-stage residual inclusion model is also applied to address endogeneity. The second-stage regression is:

[13] As discussed in greater detail in Terza, Basu and Rathouz (2008), applying two-stage least squares estimation to this type of nonlinear model may lead to inconsistent estimates. In this nonlinear setting, a residual inclusion estimation is the preferred approach.

$$Episodes_i^W = \exp(\alpha \ln(OOPP_f) + \beta_1 Z_i + \delta\widehat{\xi}_i) + e_i.$$ [14]

3.1.3 Utilization Per Episode - Intensive Margin

Analyzing the effects of the out-of-pocket price on utilization per episode may include additional information about the specific disease being analyzed. The econometric model of utilization of disease d for individual i is

$$SU_{d,i} = \exp(\alpha \ln(OOPP_f) + \beta_1 Z_i + \beta_2 X_{d,i} + \delta\xi_{i,d} + v_d) + e_{d,i},$$

where v_d is a disease-severity fixed effect and $X_{d,i}$ is a vector of covariates that includes other disease-specific information of the individual, such as an interaction between the age of the individual and the disease category. Similar to the other models, the out-of-pocket price is potentially endogenous. The potential endogeneity of the out-of-pocket price is specified in this model as the unobserved variable $\xi_{i,d}$. Again, a two-stage residual inclusion model is applied to correct for endogeneity. Unlike the previous models, a disease-specific service price index may be included. In this case, the first-stage of the estimation is

$$\ln(OOPP_f) = \gamma_1 \ln(SPI_d^r) + \gamma_2 \ln(SPI^r) + \tau_1 Z_i + \tau_2 X_{d,i} + \delta\xi_{i,d} + v_d + u_{d,i}.$$

The second-stage regression would then be:

$$SU_{d,i} = \exp(\alpha \ln(OOPP_f) + \beta_1 Z_i + \beta_2 X_{d,i} + v_d + \delta\widehat{\xi}_{i,d}) + e_{d,i}.$$

Note that there are important differences in identification when analyzing utilization along the intensive and extensive margins. When analyzing utilization along the extensive margin, an individual (or one of their family members) could potentially develop any disease, so there is a single measure of the expected service price for the entire MSA, SPI^r.[15] This limits the power for identifying demand along the extensive margin. In contrast, conditional on having a disease, the relevant service price is disease-specific. Therefore, there are many distinct disease-specific prices within an MSA that may be used to identify demand along the intensive margin.[16]

3.2 Discussion of Empirical Issues

3.2.1 Instruments

Recall the basic motivation for the instruments applied in this paper. The negotiated prices are set prior to insurance selection by consumers. Thus, the negotiated service prices will shape the incentives of insurers when offering plans, but the negotiated prices should not have a direct effect on the insurance selection

[14] In the robustness section of the paper, an additional check uses a simple count of the number of episodes.

[15] This is only approximately correct. Specifically, an alternative measure of SPI^r that is specific to each individual may be calculated based on the probability that each person develop a specific disease based on their age, sex, and other demographic characteristics. This alternative IV strategy was studied and results did not change substantially.

[16] Although some component of price variation is common across diseases, Dunn, Liebman, and Shapiro (2012) find evidence that there is a component of service price variation that is disease-specific.

by consumers. This follows the standard cost shifter argument applied in the literature.[17] A potential problem of this IV strategy is that the service prices could potentially be related to quality. For instance, higher service prices may be associated with greater quality and higher out-of-pocket prices. In this case, if individuals consume more medical care when it is of higher (lower) quality, these patterns would tend to decrease (increase) individual price responsiveness.[18] Despite this possibility, it is not clear that quality would be related to the MSA service price index. Recall that the MSA service price index is an average service price for the entire MSA, so the estimate is likely capturing a common component of costs across a large area and a variety of different providers and services, rather than the quality of a specific provider.[19] This greatly reduces the possibility of endogeneity bias. Moreover, several variables are included in the analysis to control for the quality of medical care in an MSA, such as regional fixed effects, the fraction of hospitals in the county associated with teaching facilities, and several other covariates.

To further reduce any possibility of bias, several alternative IV strategies are applied. One alternative IV strategy uses service prices from other plan types. Specifically, a service price index built from non-PPO (PPO) plans is used as an instrument for the PPO (non-PPO) out-of-pocket prices. The assumption here is that the unobserved quality is unique to a particular plan type, but common costs are shared across plan types. Both of these assumptions appear plausible. Different plan types may share common costs because they both contract with providers in the same area, but the qualities of the providers that they contract with may be different. For instance, PPO plans are likely to have a network that includes many of the highest quality physicians, while POS plans tend to be more restrictive.

As an additional check, another instrument is constructed which uses service prices in other MSAs in the state. The assumptions underlying this strategy are that unobserved demand shocks across markets are independent, but the prices are correlated due to common cost factors across MSAs in a state. The features of the market support these assumptions. Substantial evidence exists that consumer demand for medical care is local, with patients typically travelling just a few miles for inpatient services (e.g., Town and Vistnes (2001) and Gaynor and Vogt (2003)). However, labor market movements are more likely to be within states or across nearby states, creating a common cost component across a broad geographic area. Some regulations are also specific to each state (e.g., certificate of need laws for hospitals). This strategy is related to that proposed by Hausman (1996) that uses prices from other cities as an instrument for price when estimating demand.[20]

Another service price index is constructed which uses the 25th percentile of observed service prices in each MSA, rather than the average price.[21] This instrumental variable strategy may be preferred if quality

[17]As part of the insurance services offered to enrollees, insurers reduce the risk of enrollees by transforming the linear prices that they contract with suppliers, to highly nonlinear prices that reduce the risk incurred by the purchasers of insurance. One can even think of the insurance company as the retailer that is setting the out-of-pocket price of the products offered by the physician, hospital, and prescription drug wholesalers.

[18]One could imagine the bias going in either direction. Higher quality providers may have more effect on patients with a lower level of service utilization. Alternatively, higher quality treatment may involve more services.

[19]Another reason for focusing on the average service price is to avoid any potential correlation with a particular plan. While it is generally true that insurers often negotiate a single contract with providers for multiple insurance offerings, it is important that the results are robust, even if this assumption fails.

[20]It should be noted that the strategy proposed here is distinct, and perhaps less likely to be endogenous than the IV strategy applied by Hausman. The service prices in the other markets reflect the marginal cost of additional services for insurers in other MSAs. In contrast, the equivalent of Hausman's instrument in this setting would be out-of-pocket prices in other markets.

[21]Specifically, for each service category (e.g., outpatient hospital) and each disease (e.g., hypertension), the 25th percentile

differences across markets are associated with differences in price at the high end of the price distribution, but not at the low end. For example, there may be a fraction of physicians and hospitals in an area that may be perceived as very high quality (e.g., Johns Hopkins Hospital in Baltimore), while many other providers may be of more standard quality. In this case, the 25th percentile service price index may be thought of as pricing a more homogeneous medical service across MSAs.

Each of the instruments previously discussed rely on differences in service prices across MSAs. To strengthen the findings in this paper, it may be useful to find a distinct source of identification. An alternative IV strategy exploits within-MSA differences in service prices. For example, the insurer for individual i may have negotiated paying a high price for a 15-minute visit in the MSA, while the insurer for individual j in the same MSA may pay a low price. One potential IV strategy is to use these negotiated prices directly for each individual. This IV strategy assumes that prices paid to particular providers selected by the individual are unrelated to insurance selection, except through marginal cost.[22] Clearly, this assumption is very strong, since those individuals seeking the high priced physicians and hospitals may be less healthy and select more generous insurance. Therefore, to apply this strategy, but reduce the possibility of endogeneity, only the other family member expenditures are used to calculate the service price instrument. In addition, the analysis only uses an indicator of whether the family-specific service price is above or below the median price for each MSA.[23] Despite these precautions, this approach may still be problematic and will only be viewed as a robustness check that offers a unique identification strategy that may be estimated with the inclusion of MSA fixed effects.

3.2.2 Out-of-Pocket Price

The nonlinear structure of most health insurance plans makes it challenging to estimate demand, since it is unclear which price along the nonlinear schedule invokes a response by consumers.[24] For this reason, it is likely that the out-of-pocket price used in this study is only a proxy for the true out-of-pocket price. Let $OOPP_f^*$ be the out-of-pocket price perceived by the consumer, and assume that the out-of-pocket price variable used in the analysis is affected by error, v_i, so $OOPP_f = OOPP_f^* + v_i$. Much of the noise is likely created by the nonlinear nature of health insurance plans, causing v_i to shift as the amount of medical care utilization changes. The IV strategy taken in this paper helps address this problem because the negotiated price between providers and insurers is typically linear and unrelated to an individual's level of utilization, implying that $cov(v_i, \ln(SPI^r)) = 0$. In other words, the instrumental variable captures differences in the out-of-pocket price related the cost of medical services in the MSA, which is uncorrelated with the individual-specific movements in the out-of-pocket price measure.

Although the IV strategy may assist with measurement error problems, it is still necessary to select a particular measure of out-of-pocket price to include in the analysis. As described above, this paper

---

price observation is used, rather than the average.

[22] This may be the case if the different prices are due to relative negotiating clout between different insurers and providers (see Sorenson (2003)) or perhaps there are different geographic markets within an MSA.

[23] Specifically, an individual service price index using other family member expenditures is calculated as $SP_{\text{Other family members}} = \frac{\sum_{i \in \text{Other family members}} \sum_{d \in i} c_{d,i}}{\sum_{i \in \text{Other family members}} SU_i}$. The instrument is a dummy variable of whether the individual service price is above or below the median value in the MSA. The second instrument is a dummy variable that is an indicator of whether expenditures are missing for the other family members.

[24] E.g., the current price (for myopic consumers), their predicted out-of-pocket price at the end of the year (forward-looking consumers), or some average out-of-pocket price.

focuses on $OOPP_f$, which is calculated as the realized out-of-pocket expenditure for a family divided by overall utilization for the family. There are two key advantages to using this average out-of-pocket price. First, the approach does not exclude out-of-pocket payments that may be relevant. For instance, focusing on the end-of-year expected price may capture the behavior of forward-looking consumers but miss the myopic behavior of other consumers that only respond to current prices. In contrast, focusing on current prices would ignore the response of forward-looking consumers.[25] Second, the average out-of-pocket price may be easy and practical for policy-makers to apply, since it may be thought of as, roughly, the share of out-of-pocket expenditures paid by consumers.[26] For example, when applying elasticities in the literature, Newhouse (1992) and Finkelstein (2007) think about the consumer's response to out-of-pocket expenditures divided by total expenditures (i.e., an elasticity with respect to a coinsurance rate).[27] Although the paper focuses on the out-of-pocket price measure, $OOPP_f$, it is shown that the results are robust to alternative out-of-pocket price measures.

3.2.3 Empirical Model Selection

As mentioned previously, the utilization data includes skewness, heteroskedasticity, and mass points at zero, which may create statistical problems and lead the usual least squares estimation to yield bias or imprecise estimates (see Manning and Mullahy (2001)). To address these issues, a variety of statistical models and tests have been applied to determine the appropriate estimator. This analysis suggests that a GLM model with a log link and a Gamma distributional error structure fits the properties of the data nicely. This model is applied to each of the components of utilization. A discussion of the statistical tests and alternative specifications has been relegated to an appendix. However, as noted in the appendix, the key results of the paper are robust to alternative estimators, such as the application of the popular two-part model.

3.2.4 Estimating Standard Errors in a Two Stage Model

To precisely estimate standard errors of the parameters, it may be important to account for the measurement error from the first stage estimates. For this reason, a bootstrap approach is applied that repeates the two stage procedure using 50 random draws of the data with replacement. Due to the size of the data, these random draws are taken from an initial 30 percent random sample. In this particular application, it appears that the standard errors change very little when applying the bootstrap estimator, relative to estimates that ignores the impact of the first stage estimates on the second stage standard errors.

[25] It is unclear how consumers actually respond to nonlinear price schedules, so an average price is a simple way to include both myopic responses and dynamic considerations, albeit in an arbitrary fashion.

[26] The share of out-of-pocket expenditures is only roughly accurate because the measure of utilization used in the denominator will likely not equal total expenditures. Although, on average, this assumption is correct.

[27] Of course, selecting a single price to represent a nonlinearly structured insurance plan does not uncover how individuals respond to different aspects of their nonlinear insurance structure. This alternative research question is of great importance, as it may lead to a deeper understanding of consumer behavior and also determine the optimal nonlinear insurance contract (see Aron-Dine, Einav, Finkelstein, and Cullen (2012)).

4 Data

The analysis uses retrospective claims data for a sample of commercially-insured patients from the MarketScan® Research Database from Truven Health. The specific claims data used is the "Commercial Claims and Encounters Database," which contains data on medical and drug claims from employer and health plan sources for several million commercially-insured individuals, including employees, their spouses, and dependents. Each observation in the data corresponds to a line item in an "explanation of benefits" form.[28]

The sample is restricted to enrollees that are not in capitated plans from the MarketScan database for the years 2006 and 2007.[29] The sample is also limited to enrollees with drug benefits because drug purchases will not be observed for individuals without drug coverage. The MarketScan database tracks claims from all providers using a nationwide convenience sample of enrollees. Each enrollee has a unique identifier and can be linked to a particular county. All claims have been paid and adjudicated.[30]

The basic idea of looking at episodes of treatment in this paper is similar to the RAND study, but the methodology for defining and grouping episodes is distinct. In this paper, the claims data have been processed using the Symmetry grouper 7.6 software from Optum. The grouper assigns each claim to a particular Episode Treatment Group (ETG) disease category.[31] The grouper uses a proprietary algorithm, based on clinical knowledge, that is applied to the claims data to assign each record to a clinically homogenous episode of care. The episode grouper allocates all spending from individual claim records to distinct diseases.[32] An advantage of using the grouper is that it can use patients' medical history to assign diseases to drug claims, which typically do not provide a diagnosis. Another advantage is that it is replicable and the software may be applied to other data sources. Finally, the grouper algorithm is constructed by experts in the area that have a firm grasp of current diagnostic practices. However, the algorithms are also considered a "black box" in the sense that they rely entirely on the expertise of those that developed the grouper software.[33]

To ensure that all claims are properly identified and grouped into episodes, it is required that all individuals in the sample are fully enrolled for the entire year, plus 6 months prior enrollment (e.g., enrollment from July 2005 for enrollees in 2006) and 6 months post enrollment (e.g., enrollment until June 2008 for enrollees in 2007).[34] To better control for the severity of the diagnosis, additional severity

[28] The decisions made for selecting the sample and defining utilization and episodes using these data closely follow Dunn, Shapiro, and Liebman (2012).

[29] A key reason for focusing on a short cross-section is that similar medical technologies are likely available in different markets, which is an assumption that is difficult to justify when there is greater time variation.

[30] Additional details about the data and the grouper used in this paper are in Dunn et al. (2010).

[31] The ETG grouper allocates each record into one of over 500 disease groups.

[32] All episodes are initiated using only diagnostic information, so information on services or procedures performed are not used to initiate episodes. In cases where the spending could potentially be allocated to multiple diseases, the grouper uses additional information on the claim, such as the information from the patient's history or the types of procedures performed to allocate spending across disease episodes.

[33] It is worth noting that the ETG Symmetry grouper is also being applied in much of the research related to the development of a Health Care Satellite Account at the Bureau of Economic Analysis (e.g., Aizcorbe and Nestoriak (2012), Dunn, Shapiro and Liebman (2012), and Dunn, Liebman and Shapiro (2012)). They have explored applying different grouper methodologies to the data, such as the Medical Episode Grouper (MEG) from Truven Health and other ICD9 classification systems. These different grouper methodologies appear to produce qualitatively similar patterns, when looking at time trends and differences across geographic markets. Therefore, it is unlikely that the choice of disease episode grouper would have a large impact on the estimates reported in this paper.

[34] About 13.8 percent of expenditures are not assigned to any ETG disease category (that is, screening for diseases and other records that cannot be assigned a category). Those claims not assigned disease categories are removed from our analysis.

measures provided by the ETG grouper are used to further classify each episode. The availability of severity classifications vary by the ETG disease category, and range from 1 (the least severe) to 4 (the most severe). For instance, the most severe condition of diabetes will be given a severity level of 4 while the least severe diabetes condition will be given a severity level of 1.[35]

4.1 Service Utilization

Service utilization measures were created for each type of service based on the definition of a service within that service type. The service-type categories are inpatient hospital, outpatient hospital, general physician, specialist physician, prescription drug, and other. Measuring service utilization is not a straightforward task since the definition of a "service" is a bit ambiguous and there are a variety of ways that one could define it across various service types. Ideally, the definition of a specific service price should depend on how the price of that service is typically set and paid. For example, for physician services, insurers contract to pay a unique price for each procedure (that is, the insurer and the patient together pay this amount), whereas the prices paid to facilities are often set based on the treated disease. The next subsection describes how the quantity of services is measured for each service type.

4.1.1 Measuring the Quantity of Service by Service Type

Physician office - Physician visits are based on procedures performed in a physician's office. Since not all procedures are equivalent, each procedure is assigned a weight by a variable similar to a "Relative Value Unit" or "RVU", which measures the service intensity of each procedure and is used by Medicare to reimburse physicians for each procedure that is performed.[36] Specifically, for each current procedural terminology (CPT) code and modifier code, the average fee for that procedure performed in an office setting across all locations is calculated. The total quantity of services performed in an office is then computed by summing over these calculated RVU units. Drawing on the previous discussion, one can think of the RVU as a base price used to compute a utilization index, similar to the utilization amount calculated in equation (3). More precisely, the total amount of RVU units from an office visit is $\sum_{cpt \in Visit} \overline{p}_{cpt}$, where $cpt \in Visit$ is a complete list of CPTs performed during the visit and $\overline{p}_{cpt}$ is the price for the CPT code for the base group, where the base group price is the average price in the data. For instance, if a CPT code indicating a 15-minute office visit has an average price of \$100 across the data, its value will be 100 RVUs. It should be clear that this sum of RVU units is a measure of utilization and not price. To see this, note that if one observes that the fee on a 15-minute office visit is \$120 in an area, then the price of the service will be calculated as \$120/100RVU=1.2 \$/RVU.[37]

As mentioned in the robustness check section in the appendix, the main results do not change when these ungrouped claims are incorporated into the analysis.

The six-month "cushion" ensures that episodes occurring at the beginning or the end of a year are not truncated. The results do not appear sensitive to this six-month cushion.

[35] The ETG severity level is determined for each episode based on a variety of additional information including age, gender, comorbidities, and other potential complications.

[36] This framework has also been adopted by the commercial market. In a survey of 20 health plans conducted by Dyckman & Associates, all 20 health plan fee schedules were influenced by a resource-based relative value scale (RBRVS), which is a pricing methodology that applies RVUs. Taking the average of observed prices in the market for each procedure is one measure used for capturing the typical "resources" used for a procedure.

[37] This methodology for calculating utilization for physician services is identical to that conducted by Dunn and Shapiro (2012).

Hospital inpatient - Inpatient hospital stays not only consist of facility fees paid to the hospital, but also fees paid to the physician. For the portion of fees paid to the hospital, the amount of services is measured as the average dollar amount for an inpatient stay for the observed disease. Again, this average dollar amount for an inpatient stay may be viewed as the price for the base group in a utilization index. For the portion of fees paid to the physician, an RVU is assigned in the same way as in the office setting. However, procedure price averages are taken only from the inpatient setting. The total amount of services performed in an inpatient setting is calculated by adding the physician and facility amounts.

Hospital outpatient - Outpatient hospital visits are calculated in an identical fashion to the inpatient hospital visits. That is, the facility amount is calculated based on the average outpatient visit for that disease, and the doctor's portion of the total amount is calculated based on the average payment for the procedure codes in the outpatient setting.

Prescription drugs - The amount of the prescription drug varies based on the molecule, the number of pills in the bottle, the strength of the drug, and the manufacturer. Each combination of these factors is assigned a unique NDC code.[38] To capture these differences, the average price for each NDC code is calculated. This means that branded and generic products that contain the same active molecule are treated as distinct drugs. The average price for each NDC code represents the amount of the service used. If the expenditure on a prescription is greater than this amount, it suggests that prices are above average in an area.

All other - The other category primarily includes ambulatory care, independent labs, and emergency room visits. For these services, the amount of each category is measured as the average cost for a visit to that particular place of service (for example, the average cost of an ambulatory care visit to treat ischemic heart disease). For cases where procedure codes are available, the average cost of that procedure code for that place of service is applied.

These measures of service quantity subsequently allow for the creation of a service price that corresponds well with how fees are negotiated in the marketplace. In practice it appears that physicians and hospitals often negotiate on a percentage amount from some pre-determined base, such as 10 percent above Medicare rates.[39] As the measure of service price can be intuited as expenditure divided by a proxy for "RVUs", it can also be thought of as a percentage amount from a base (or average) payment—a measure close to how prices are actually set.

Additional details regarding the measurement of service quantity are discussed in Dunn, Shapiro, and Liebman (2012) and the corresponding technical appendix to that paper.

4.2 Sample and Descriptive Statistics

The sample studied in this paper is limited to those MSAs with a sufficiently large number of enrollees, so that the measured service prices in each market will be meaningful. The sample includes only those MSAs in the data that have an average of 15,000 enrollees per year over the 2006-2007 time period.[40] The minimum sample size in each city is more than double the annual commercially-insured sample size from the Medical Expenditure Panel Survey, which is a national survey of health expenditures meant to

[38] An 11-digit National Drug Code (NDC) uniquely identifies the manufacturer, the strength, dosage, formulation, package size, and type of package.

[39] As mentioned previously, of the 20 plans surveyed by Dyckman & Associates (2003), all of the plans use some variation of the Medicare resource-based relative value scale (RBRVSD) methodology to set prices.

[40] i.e., 30,000 enrollee-year observations.

be representative of the entire U.S. non-institutionalized population.[41] This first selection rule leaves a sample of 103 MSAs.

All disease episodes are considered when studying the effects of out-of-pocket price on utilization. However, when constructing the MSA service price indexes (i.e., SPI^r) to use as instruments, only those diseases that have 15,000 episodes or more in the data are selected, which accounts for 87 percent of overall expenditures and 96 percent of the episodes. The reason for this selection rule is to make sure that the price indexes are not greatly affected by infrequently observed diseases.

Table 1 provides some basic descriptive statistics for the top spending disease categories. Prior to calculating these descriptive statistics, population weights are applied to adjust for differences in age and sex across MSAs and to make the estimates representative of U.S. totals.[42] The table reports the national estimates of expenditures for each disease along with the number of episodes, dollars per episode, and expenditure share. The table reveals some interesting facts about disease expenditures in these data. First, based on the ETG groupings, the top five disease expenditure categories include pregnancy, joint degeneration of the back, hypertension, diabetes, and ischemic heart diseases. Although there are 271 disease-severity combinations in the sample, these five disease categories account for 25 percent of the expenditures. In general, most of the expenditures are accounted for by a limited number of diseases with the diseases listed here accounting for 38 percent of total expenditures from the selected diseases, so the MSA service price indexes will be heavily influenced by a small number of diseases. There is a wide range in the expenditure per episode across diseases. Severity 1 hypertension costs just \$646 per episode, while severity 3 joint degeneration of the back costs \$12,555.

[41] The commercially-insured sample in the MEPS data is around 14,799 individual observations in each year. This study uses two years of data which includes more than 30,000 individual-year observations per MSA. The sample size of MSAs is larger than that used in Dunn, Shaprio, and Liebman (2012). Similar results are obtained with a smaller sample of MSAs, but more cities ensures that the estimates are representative.

[42] Specifically, enrollees in each MSA are assigned weights so the weighted population has an age and sex distribution that is identical to that of the U.S. commercially-insured population in 2007. For constructing the MSA service price indexes, population weights are also applied to each MSA so that the service price estimates are unaffected by the demographics of the population. See Dunn, Liebman and Shaprio (2012) for more detail.

Table 1 shows the disease expenditures for the two-year period of 2006 and 2007 and is based on the weighted sample of enrollees. The national weights are applied to each city and the total expenditures and episodes are divided by the number of cities in the sample, 103, times the number of years of data, 2. (Thus I divide by 206 (=103*2)). Since these figures do not account for differences in populations across cities, these estimates overcount smaller MSAs, relative to their share of the U.S. population.

Table 1. Summary Statisics on Top Spending Disease Episodes

	Disease	Severity	Total Dollars (Billions) 2006-07	Number of Episodes (Thousands)	Dollars Per Episode	Share of Spending
1	Pregnancy, with delivery	1	$7.0	1,493	$9,377	3.4%
	Pregnancy, with delivery	2	$3.5	505	$13,834	1.7%
2	Joint degeneration, localized - back	1	$5.6	6,313	$1,774	2.7%
	Joint degeneration, localized - back	2	$2.5	1,174	$4,270	1.2%
	Joint degeneration, localized - back	3	$1.9	302	$12,555	0.9%
3	Hypertension	1	$5.7	17,638	$646	2.7%
	Hypertension	2	$1.7	3,836	$868	0.8%
	Hypertension	3	$0.9	1,573	$1,090	0.4%
	Hypertension	4	$0.7	584	$2,240	0.3%
3	Diabetes	1	$5.0	6,428	$1,543	2.4%
	Diabetes	2	$0.9	716	$2,464	0.4%
	Diabetes	3	$0.9	535	$3,287	0.4%
	Diabetes	4	$1.4	469	$5,915	0.7%
4	Ischemic heart disease	1	$4.3	2,373	$3,631	2.1%
	Ischemic heart disease	2	$3.3	1,195	$5,588	1.6%
5	Routine exam	1	$6.7	62,047	$215	3.2%
6	Mood disorder, depressed	1	$4.3	7,208	$1,184	2.0%
	Mood disorder, depressed	2	$0.9	1,152	$1,575	0.4%
	Mood disorder, depressed	3	$0.6	357	$3,130	0.3%
7	Hyperlipidemia, other	1	$5.2	15,989	$649	2.5%
8	Joint degeneration, localized - neck	1	$3.2	4,160	$1,519	1.5%
	Joint degeneration, localized - neck	2	$0.5	372	$2,427	0.2%
	Joint degeneration, localized - neck	3	$1.3	292	$9,168	0.6%
9	Chronic sinusitis	1	$2.7	10,345	$512	1.3%
	Chronic sinusitis	2	$0.6	1,336	$890	0.3%
	Chronic sinusitis	3	$1.0	888	$2,332	0.5%
10	Asthma	1	$1.2	4,231	$575	0.6%
	Asthma	2	$1.7	3,431	$980	0.8%
	Asthma	3	$0.3	336	$1,918	0.2%
	Asthma	4	$0.6	288	$3,917	0.3%
	Other		$133	373,029	$711	63.6%
	Total		$208	530,598	$785	100.0%

Table 2 provides descriptive statistics on many of the variables used in the analysis at the individual level. The table shows that the majority of the data is from large employers, with only 24 percent of the sample contributed by insurers. The data is also comprised mostly of enrollees in PPO plans, accounting for 68 percent of the sample.[43] Variables from external data source are also incorporated into the analysis to control for factors that may affect medical utilization that are not contained in the MarketScan data. One data sources is the Area Resource File (ARF) database that includes several county-level variables, such as the median income, fraction of individuals with a college education, average rent,[44] and the fraction of hospitals associated with a medical school in the county. Another external data source is the Behavioral Risk Factor Surveillance System (BRFSS) data that is used to construct measures of health, including estimated rates of obesity and smoking in each county.[45]

[43] This compares with 60 percent reported in the Kaiser Health Benefit Survey in 2006. Although the share of PPOs may appear high, recall that all capitated plans, such as HMOs, have been dropped from the analysis. Taking into account those HMO enrollees would produce estimates very similar to the Kaiser Health Benefit Survey.

[44] Although the average rent would not affect medical care utilization directly, it may be related to the price of outside goods and services in the area.

[45] One limitation of these supplementary variables is that they do not include individual-specific information, but only county-wide information. However, the inclusion of these additional variables ensures that the relationship between price and utilization across areas is not driven by these county-specific factors.

The estimates from the BRFSS data are based on regression analysis at the individual level that are used to compute county-level estimates. To standardize the estimates, rates of obesity and smoking are computed for a standardized individual in the county (i.e., a woman of age 34 to 44). Unfotunately, the BRFSS data only includes an indicator of whether a person has insurance, and does not include information regarding the source of their coverage, such as Medicaid or employer-based

Table 2 also shows measures of utilization and price. Note that each of the utilization measures are highly variable and around 16.5 percent of enrollees consume zero health services. For those that do consume a positive amount of health services, the mean utilization amount is 3,967 and the standard deviation is 10,783. The utilization is also highly skewed to the right, as can be seen by comparing the mean to the median. To address the skewness of the data the demand analysis will focus on log transformations of the utilization measures.

The bottom of the table reports the various price measures. One striking feature of the data is that the variation in the out-of-pocket price variable, $OOPP_f$, is extremely large, with a coefficient of variation of about 1. This measure of variation is much larger than the variation in the other price measures. The MSA service price index, SPI^r, has a coefficient of variation of 0.086, and the disease-specific service price index, SPI^r_d, has a coefficient of variation of 0.147. Although the variation on $OOPP_f$ appears large, this should be expected, since the out-of-pocket price is specific to each individual and is affected by the various nonlinear characteristics of the insurance contracts. In contrast, the service prices negotiated between insurers and providers are typically linear. The variation in the service price indexes is also smaller because it averages over prices for the entire MSA, eliminating differences in contracted amounts within an MSA. The considerable noise contained in the $OOPP_f$ variable implies that a substantial amount of variation in the MSA service price index may be necessary to accurately identify the relationship between the service price index and the out-of-pocket price. Fortunately, there are clear differences in the MSA service price indexes across areas, ranging from 0.89 to 1.10 for the 10th and 90th percentiles. This observed variation in the service price index is critical for the successful application of the IV strategy applied in this paper.[46]

insurance. Prior to estimating the regression model, those individuals that do not have insurance and also those households that earn less than $10,000 annually are removed from the analysis. Those without insurance clearly do not match with our population of commercially-insured individuals. In addition, households that earn less than $10,000 are much more likely to be enrolled in Medicaid or another public assistance program and not be included in the commercially-insured population. Additional health factors, such variables related to drinking, exercise or BMI, were included in the analysis, but had no effect and potentially introduce multicollinearity with the other county-level health variables.

[46] See Dunn, Shapiro, and Liebman (2012) for a more complete discussion and analysis of service price variation.

Table 2. Descriptive Statistics

	Mean	Median	s.d.	10th percentile	90th percentile
Age	33.324	37.000	19.816	4.000	58.000
Number of Individuals in the Family	2.796	3.000	1.507	1.000	5.000
Fraction with College Education (in County)	0.166	0.155	0.062	0.094	0.255
Income (Median in County)	$56,607	$53,472	$13,832	$41,845	$75,460
Rent (Median in County)	$647	$633	$133	$492	$835
Fraction of Hosp. Med. Schools (in County)	0.387	0.368	0.319	0.000	0.875
Fraction Obese (in County)	0.236	0.237	0.060	0.162	0.316
Fraction Smokers (in County)	0.170	0.167	0.049	0.111	0.230
Male	0.486				
Data Source: Insurer Data	0.239				
Plan Type					
PPO	0.681				
POS	0.165				
Comprehensive	0.062				
High Deductible Health Plan	0.034				
EPO & Other	0.058				
Overall Service Utilization (SU_i)					
SU_i=0	0.1646				
SU_i if SU_i>0	3967.39	1271.19	10783.99	176.48	9009.45
Number of Episodes					
Simple Count ($Episodes_i$) if SU_i>0	3.68	3.00	2.50	1.00	7.00
Weighted Count ($Episodes^w_i$) if SU_i>0	3981.91	2145.56	5396.77	389.22	9822.67
Service Utilzation Per Episode ($SU_{i,d}$)					
$SU_{i,d}$	740.22	196.31	2433.62	57.70	1519.91
Out-of-pocket Price and Service Price Variables					
$OOPP_f$	0.232	0.186	0.228	0.056	0.442
MSA Service Price (SPI^r)	1.000	0.998	0.086	0.894	1.104
Disease-specific, MSA Service Price (SPI^r_d)	1.002	0.993	0.147	0.847	1.158
Number of Individuals	9,735,083				
Number of Episodes	32,592,524				

Notes: The data sources for the individual-level variables are from MarketScan. The county level variables are from the ARF and BRFSS data sources and are linked to the individual observations through the observed county of the individual in the MarketScan data. The total number of individuals and episode observations are reported at the bottom of the table. The total observations do not match the totals reported in the estimates, since not all the variables are observed for all individuals.

Table 2 reports nearly 10 million individuals in the sample, but not all of the variables are observed for every individual in the data, so a more limited sample is used for estimation. For the main estimates, $OOPP_f$, is imputed for families with zero expenditures using information from similar families in the same MSA (approximately 8 percent of the individual observations),[47] although the results remain unchanged when the imputed observations are removed.[48] The next section presents the main empirical findings, which show the effects of out-of-pocket price on each of the measures of utilization.

[47] Specifically, individuals from families of the same size, age, sex, plan-type, and data contributor (employer or insurer) in the same MSA are used for imputation. To conduct the imputation, total out-of-pocket expenditures and total utilization are calculated for each demographic category. Then the $OOPP_f$ is imputed by dividing total out-of-pocket expenditures by total utilization for individuals of the same category.

[48] These estimates are shown in the robustness section in the appendix.

5 Results

5.1 Overall Utilization

Table 3 presents estimates of the overall utilization response to the out-of-pocket price. All of the estimates include the controls listed in Table 2 along with regional fixed effects, although only selected parameter estimates are displayed.[49] Model 1 shows the baseline results that do not control for endogeneity. The price elasticity implied by Model 1 is -0.62, which is considerably more elastic than most other estimates in the literature, suggesting a negative bias. Indeed, evidence of a negative bias is found by looking at the estimates of Model 2 that applies the MSA service price index instrument. For Model 2, the estimates show a price coefficient of -0.22, which is considerably more inelastic than the estimates from Model 1. Moreover, the coefficient on the residual inclusion variable (derived from the first-stage of the estimation routine) is negative and highly significant, indicating that controlling for endogeneity is statistically important and endogeneity bias is likely affecting Model 1 estimates. The estimates in Model 2 are only slightly more elastic than those reported in the RAND study, which centered around -0.20.[50]

Table 3. Effect of Out-of-pocket Price on Overall Service Utilization (SU_i)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Log($OOPP_f$)	-0.620***	-0.223***	-0.322***	-0.161***	-0.199***	-0.276***	-0.312***
	(-40.33)	(-3.47)	(-3.99)	(-2.76)	(-3.75)	(-2.76)	(-9.34)
Log(Median Income)	0.126	0.189***	0.173***	0.205***	0.198***	0.184***	0.246***
	(1.63)	(3.57)	(3.42)	(4.87)	(4.97)	(3.33)	(4.76)
Log(Frac. Obese)	-0.0908**	-0.0346	-0.0472**	-0.0172	-0.0231	-0.0425*	-0.00386
	(-2.11)	(-1.50)	(-2.06)	(-0.70)	(-0.95)	(-1.67)	(-0.17)
Log(Frac. Smokers)	0.0123	0.0214	0.0191	0.0270	0.0260	0.0200	0.0318
	(0.46)	(1.35)	(1.33)	(1.62)	(1.53)	(1.28)	(2.11)
Log(Frac. w/ College)	0.0301	0.0368	0.0357	0.0296	0.0287	0.0351	0.0526
	(0.58)	(1.27)	(1.18)	(1.28)	(1.13)	(1.15)	(1.89)
Residual Inclusion		-0.408***	-0.303***	-0.464***	-0.426***	-0.348***	-0.336***
		(-7.01)	(-3.85)	(-8.64)	(-9.34)	(-3.71)	(-10.73)
Number of Observations	8979207	8979207	8979207	8079984	8079984	8979207	8979207
Instruments	None	MSA Service Price	MSA Service Price Other Plans	Service Price of Other MSAs in State	Service Price of Other Plans & Other MSA Price	MSA Service Price, 25th Percentile	Within MSA Service Price Instruments

Notes: The z-statistics are in parentheses and are clustered by MSA. The z-statistics are computed using a bootstrap estimation that accounts for the two-stage estimation strategy. One, two, and three asterisks indicate significance at the 10-percent, 5-percent, or 1-percent significance level, respectively. The coefficients on the other explanatory variables are shown in Table A1.1 for select models.

As discussed previously, one potential concern with the instrument applied in Model 2 is that the MSA service price index, $\ln(SPI^r)$, may be correlated with unobserved quality. To address this potential problem, several alternative estimates are presented that apply distinct IV strategies. Model 3 uses an

[49] Complete parameter estimates of selected specifications are shown in Table A1 of the appendix.

[50] The full set of estimates for Models 2 is included in the appendix in Table A1.

MSA service price index constructed from other types of health plans in the area; and Model 4 uses a service price index constructed from other MSAs in the state. The price elasticity estimate from Model 3 is a bit more elastic than Model 2, with an elasticity of -0.32; while the estimate in Model 4 is less elastic, showing a coefficient of -0.16. Model 5 presents the preferred IV strategy, which uses both of the instruments from Models 3 and 4. Model 5 strategy is preferred since both of the instruments do not rely directly on price information from the enrollee's plan, both of the instruments contribute significantly to explaining the variation in out-of-pocket prices, and both pass basic tests of validity, discussed later. This preferred IV strategy appears to produce estimates that are practically identical to those in the RAND study.

Although Model 5 is the preferred specification, two additional estimates offer some additional checks. The instrument used in Model 6 is identical to that of Model 2, except instead of using an instrument based on average prices, Model 6 is based on the 25th percentile price for every disease and service category.[51] This strategy is distinct, but one can see the estimated elasticity in Model 6 is comparable to the other IV estimates.

Finally, in contrast to Models 2 through 6 that rely on across-MSA variation in prices to construct instruments, Model 7 exploits within-MSA variation in service prices. The Model 7 instrument only indicates whether the negotiated service prices paid by the other family members are above or below the median in the MSA. To ensure that effects are identified based on only within-MSA variation in prices, MSA fixed effects are included in this specification. The results in Model 7 are very similar to the previous estimates.[52]

In addition to the coefficient on the out-of-pocket price variable, Table 3 reports a number of other estimates of potential interest. One finding is that the measure of county obesity rates and smoking rates are unrelated to overall utilization. This result is surprising given that obesity and smoking are related to the development of particular diseases. This may suggest that these populations may not be seeking treatment for existing medical conditions, although studying these data at the disease level shows that higher obesity rates are significantly related to treatment for diabetes and hypertension.[53] Another issue is that these variables are only proxies for obesity and smoking for an entire county, rather than the precise measurement for an individual person. The coefficient on household median income is positive and highly significant, as expected. The coefficient is around 0.21, which is close in value to estimates from the RAND study, although a bit on the high end. The RAND study suggests an income elasticity of demand of 0.20 or less (see Phelps (1992)).[54] One possible reason that the elasticity is slightly larger is that this is an estimate of the overall elasticity, and the RAND study focused on the elasticity along the extensive margin (i.e., the number of episodes) rather than overall utilization. Estimating demand along the extensive margin is the topic of the following subsection.[55]

[51] For example, the 25th percentile in prices for services in a physician office to treat hypertension.

[52] While this approach offers an important check on previous results, recall that this instrumental variable strategy should be viewed as more problematic than Models 2 through 6, since negotiated prices may be directly related to plan selection (e.g., those that choose the high-priced physicians, may also choose the plan with the most generous benefits). Including or excluding the MSA fixed effects from the estimation had no effect on this result.

[53] This is observed for the disease-specific estimates in Table A2 of the appendix.

[54] The calculations reported in Phelps (1992) are derived from the estimates from Keeler et al. (1988).

[55] This elasticity is likely capturing only the demand response of the consumer, and not the larger "general equilibrium" income response that includes the effect of income on the adoption of new technologies, which may be considerably larger. Acemoglu, Finkelstein and Notowidigdo (2013) estimate a general equilibrium income elasticity of 0.7.

5.2 Extensive Margin: Weighted Number of Episodes

Table 4 examines the effect of out-of-pocket price on the weighted number of episodes.[56] For nearly all of the key estimates, Models 2 through 6, the price responsiveness matches the results found in Table 3. This confirms a key finding from the RAND study: consumers primarily respond to out-of-pocket prices by changing the number of episodes treated, rather than the utilization per episode. That is, this finding is consistent with a simple model of consumer behavior where individuals choose whether to be treated for a disease episode or not, but have less control over subsequent utilization.

Overall, the elasticities are quite close to those of the RAND study with the key result from Model 5 matching the RAND elasticity. Another interesting finding in Table 4 is that the income elasticity ranges from 0.10 to 0.20, a range that is comparable with the estimates reported by Phelps (1992). The one estimate in Table 4 that stands out as distinct from the rest is Model 7, which shows that the elasticity estimate is around -0.13. Although this estimate is markedly less elastic than the other estimates, the key qualitative finding still holds: price responsiveness is negative, significant, and inelastic.

Table 4. Effect of Out-of-pocket Price on Weighted Number of Episodes (Episodes$^{w}_{i}$)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Log(OOPP$_f$)	-0.295***	-0.228***	-0.276***	-0.181**	-0.197***	-0.269***	-0.130***
	(-36.58)	(-3.91)	(-4.24)	(-2.45)	(-3.19)	(-3.07)	(-4.68)
Log(Median Income)	0.106**	0.118***	0.110***	0.130***	0.127***	0.111***	0.188***
	(2.52)	(3.70)	(3.13)	(3.31)	(3.31)	(2.97)	(6.71)
Log(Frac. Obese)	-0.0246	-0.0151	-0.0218	0.00105	-0.00153	-0.0209	0.00852
	(-1.02)	(-0.65)	(-0.92)	(0.05)	(-0.07)	(-0.82)	(0.37)
Log(Frac. Smokers)	-0.00503	-0.00345	-0.00458	-0.00201	-0.00247	-0.00444	0.0125
	(-0.37)	(-0.26)	(-0.32)	(-0.16)	(-0.20)	(-0.33)	(0.95)
Log(Frac. w/ College)	-0.0318	-0.0306	-0.0314	-0.0363	-0.0364	-0.0313	0.00235
	(-1.23)	(-1.32)	(-1.31)	(-1.56)	(-1.51)	(-1.26)	(0.11)
Residual Inclusion		-0.0693	-0.0201	-0.115*	-0.0986*	-0.0267	-0.177***
		(-1.24)	(-0.32)	(-1.63)	(-1.71)	(-0.31)	(-7.25)
Number of Observations	8979207	8979207	8979207	8079984	8079984	8979207	8979207
Instruments	None	MSA Service Price	MSA Service Price Other Plans	Service Price of Other MSAs in State	Service Price of Other Plans & Other MSA Price	MSA Service Price, 25th Percentile	Within MSA Service Price Instruments

Notes: The z-statistics are in parentheses and are clustered by MSA. The z-statistics are computed using a bootstrap estimation that accounts for the two-stage estimation strategy. One, two, and three asterisks indicate significance at the 10-percent, 5-percent, or 1-percent significance level, respectively. The coefficients on the other explanatory variables are shown in Table A1.1 for select models.

[56] One useful by-product of modeling demand in this manner is that the data on weighted number of episodes is much less skewed than overall utilization, as shown in Table 2.

5.3 Intensive Margin: Utilization Per Episode

To complete the picture of demand responsiveness, Table 5 reports estimates for the effect of out-of-pocket price on the utilization per episode.[57] Model 1 of Table 5 shows the results of the baseline model that does not control for endogeneity. The estimates show the relationship between out-of-pocket price and utilization to be negative and highly statistically significant. This estimate is dramatically different than the results of Models 2 through 7 that each correct for endogeneity and show price elasticities that are more inelastic and less statistically significant. The key estimate from Model 5 suggests an elasticity of around -0.05. This result is in line with expectations based on the estimates from the previous two tables. That is, the intensive margin elasticity should be roughly equal to the overall utilization elasticity (Table 3) minus the extensive margin elasticity (Table 4). Also, as one might expect, the estimates from Model 7 are distinct, showing a negative and significant elasticity. Again, although the estimate from Model 7 is different, the result is qualitatively similar to the other estimates. Specifically, price responsiveness is highly inelastic and controlling for endogeneity substantially reduces the magnitude of the elasticity.

Table 5. Effect of Out-of-pocket Price on Utilization per Episode ($SU_{d,i}$)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Log(OOPP$_{i}$)	-0.196***	-0.0252	-0.0494*	-0.0672*	-0.0503*	-0.0436*	-0.103***
	(-40.84)	(-1.29)	(-1.76)	(-1.80)	(-1.80)	(-1.67)	(-14.53)
Log(Median Income)	-0.0335*	0.00644	-0.00963	-0.00411	-0.00937	0.00300	0.0251*
	(-1.80)	(0.34)	(-0.51)	(-0.21)	(-0.48)	(0.15)	(1.70)
Log(Frac. Obese)	-0.0650***	-0.0374***	-0.0444***	-0.0495***	-0.0492***	-0.0405***	-0.0146***
	(-6.65)	(-4.15)	(-4.51)	(-4.91)	(-5.02)	(-3.98)	(-3.28)
Log(Frac. Smokers)	0.0134**	0.0184***	0.0164**	0.0197***	0.0187**	0.0180**	0.0118**
	(1.97)	(2.67)	(2.36)	(2.73)	(2.56)	(2.51)	(2.53)
Log(Frac. w/ College)	0.0288***	0.0304***	0.0268**	0.0293***	0.0252**	0.0298***	0.0270***
	(2.72)	(3.06)	(2.51)	(2.61)	(2.21)	(2.80)	(3.12)
Residual Inclusion		-0.177***	-0.128***	-0.127***	-0.125***	-0.154***	-0.107***
		(-9.25)	(-4.71)	(-3.40)	(-4.59)	(-5.99)	(-15.30)
Number of Observations	28533369	27812331	23813449	25835792	21561379	27812331	28533369
Instruments	None	MSA Service Price	MSA Service Price Other Plans	Service Price of Other MSAs in State	Service Price of Other Plans & Other MSA Price	MSA Service Price, 25th Percentile	Within MSA Service Price Instruments

Notes: The z-statistics are in parentheses and are clustered by MSA and Major Practice Category. Due to the larger number of observations, the z-statistics are not adjusted for the two-stage estimation. However, applying a boostrap estimate that accounts for the two-stage estimation produces z-stats slightly larger than those reported in Model 5. One, two, and three asterisks indicate significance at the 10-percent, 5-percent, or 1-percent significance level, respectively. The coefficients on the other explanatory variables are shown in Table A1.2 for select models.

Table 5 contains a few additional estimates of interest. First, income has little effect on utilization along the intensive margin, but the fraction of individuals with college education appears to have a positive and significant effect on utilization. One possible explanation may be that more highly educated individuals

[57] These estimates only focus on those diseases that are observed 15,000 times or more in the data to eliminate influence of costly and rare disease episodes.

tend to comply with prescribed treatments.[58] Second, a higher fraction of smokers in an area is associated with higher levels of utilization, perhaps due to some unobserved illness severity for this population. Surprisingly, higher obesity rates are associated with less utilization per episode. This result is unexpected, but one possible explanation is that the possible stigma associated with obesity may lead obese enrollees to avoid recommended medical care.[59]

Overall the estimates in Tables 3, 4, and 5 show patterns that are consistent with the RAND study. The similarity of the findings may be seen as somewhat surprising given the dramatic changes in health care markets in the past decades and the different approach taken to estimating demand. On the other hand, these results may simply suggest that this alternative methodology offers a reasonable and accurate approach for identifying demand elasticities and the behavior of consumers has not markedly changed since the RAND study. The following subsections further investigate the robustness of these results.

5.4 Empirical Issues and Robustness Checks

The empirical strategy in this paper rests on the strength and validity of the instruments. The first-stage estimates are reported in Table 6. The relationship between each of the service price indexes and the out-of-pocket price is strong and statistically significant. Looking at Model 4 in Table 6 shows that two of the key instruments, the average price from other MSAs in the state and the average price for other plan types in the MSA, are both significant, even when jointly included in the regression. This indicates that these instruments explain distinct components of out-of-pocket price variation. Interestingly, the first-stage coefficient on the MSA service price index is around 2, suggesting that a disproportionate share of the service price is passed on to consumers through higher out-of-pocket prices. This coefficient implies that a 10 percent increase in the service price index in an MSA tends to be associated with a 20 percent increase in the out-of-pocket price. Similar magnitudes are found using the other instruments. This is not only an important descriptive statistic, but it also has clear implications for antitrust authorities attempting to link price changes from anticompetitive provider mergers to consumer harm. Specifically, these results suggest that higher negotiated prices are likely to harm consumers as prices are disproportionately passed onto consumers through higher out-of-pocket prices.

[58] Goldman and Smith (2002) report that more educated people are more likely to comply with diabetes and AIDS treatment, conditions considered highly demanding for proper compliance.

[59] Sundmacher (2012) shows that obese individuals do not change health related behavior after a negative change in their own health (i.e., health shock). In contrast, the study finds that smokers do change their behavior. Louis and Drury (2002) find that a higher body mass index is associated with an increase in the delay or avoidance of health care.

Table 6. First-Stage Estimation of Log($OOPP_f$) on Service Price Instruments

	(1)	(2)	(3)	(4)	(5)
Log(MSA Service Price)	2.061***				
	(8.73)				
Log(MSA Service Price, Other Plans)		1.654***		0.930***	
		(6.33)		(3.22)	
Log(Service Price of Other MSAs in State)			2.265***	1.729***	
			(7.80)	(4.90)	
Log(MSA Service Price, 25th Percentile)					1.499***
					(5.73)
Number of Observations	8979207	8979207	8079984	8079984	8979207

Notes: The z-statistics are in parentheses and are clustered by MSA. The table only displays the coefficients on the insturments, with the other first-stage estimated coeffients not shown. One, two, and three asterisks indicate significance at the 10-percent, 5-percent, or 1-percent significance level, respectively.

In addition to an instrument being strong, a good instrument should be uncorrelated with the error (i.e., for an unbiased estimate, $\ln(SPI^r)$ must be uncorrelated with ξ_i). Several checks are conducted to determine whether this criteria is satisfied. One informal check is to note that a variety of distinct IV strategies produce elasticities in a reasonably tight range, from -0.16 to -0.32 for the overall utilization elasticity. The similarity across different IV strategies implies that these distinct instruments are capturing the same economic effect and that the observed differences are likely driven by sampling error. As an additional check of the validity of the instruments, the residual from the key estimates (Model 5 in Tables 3) is regressed on all of the exogenous variables, including the two instruments. A joint F-test of these two instruments shows that they are statistically insignificant, and the R-squared is extremely low 0.0029, suggesting little correlation between the instruments and the error term.[60] Checks are also performed on the extensive and intensive margin with similar results.[61]

While basic statistical checks suggest that the instruments are appropriate, another useful exercise is to examine out-of-pocket price responsiveness for specific diseases.[62] To check the responsiveness at the disease level, the analysis focuses on relatively common diseases that do not always require immediate

[60] Additional checks are conducted by looking the relation between the MSA service price index (i.e., the instrument in Model (1) of Table 6) and the residuals from the demand estimates using the two instruments (i.e., the instruments applied in Model (5) of Table 3). Again, there is no significant correlation. Similarly, there is no correlation between the residual and the MSA Service Price, 25 Percentile instrument.

The MSA service price index is not used in combination with the other two instruments because the explanatory power of the MSA-specific instrument dominates the other instruments. Morever, the theoretical possibility of endogeneity using the MSA service price index is greater than the other instruments, implying that its greater explanatory power may be problematic.

[61] The extensive margin check for endogeneity closely matches the results of the overall utilization test. For the intensive margin calculations, using the disease-specific prices as instruments produces an F-test that is statistically significant, suggesting that there may be some correlation with the error term. However, when the disease-specific instruments are not used, the out-of-pocket price coefficient remains very inelastic and of a similar magnitude. To improve the efficiency of the estimates, the disease-specific prices are included in the estimation.

[62] One must be cautious in conducting this type of analysis, since the diagnosis and treatment path may be highly nonlinear. For instance, for more serious conditions, such as for heart disease, proper treatment of high cholesterol may influence the probability of heart disease appearing in the population. Therefore, a lower price in an area, leading to more high cholesterol treatment, may actually lower the probability of seeking treatment for heart disease. When analyzing the price responsiveness for specific diseases, the implicit assumption is that they are not affected by these types of nonlinearities.

treatment, such as high cholesterol, diabetes, hypertension, migraines and depression. These estimates are contrasted with the effect of out-of-pocket price on the probability of treatment for appendicitis, which is a condition that arguably afflicts people more randomly and must be treated regardless of the price.[63] Therefore, the price responsiveness for treating appendicitis offers a falsification exercise. Instrumental variable probit models are estimated to examine if treatment for these diseases is sensitive to the out-of-pocket price. The estimates are shown in Table A2 of the appendix. As expected, the probability of being treated for high cholesterol, hypertension, diabetes, migraines and depression are each negatively related to the out-of-pocket price (though insignificant for high cholesterol). In contrast, there appears to be no significant relationship between appendicitis and out-of-pocket price, as expected.

Dartmouth researchers and others studying geoographic variation have documented significant variation in how physicians practice medicine across geographic markets (see Skinner (2012). Therefore, one may also be concerned that the observed prices may be related to physician norms and practices across markets. For instance, it may be that high price areas are those areas where physician utilization tends to be low, causing the observed negative correlation between price and utilization. To control for this possibility, a robustness check is included in the appendix that accounts for the propensity of physicians to utilize medical services by including the log of Medicare expenditures per capita. Per capita Medicare expenditures should reflect the propensity to utilize services, since the same physicians often treat both Medicare and commercial patients. Moreover, Medicare service prices are relatively flat across markets, so expenditures per capita primarily reflect utilization differences (see Gottlieb et al. 2010). The robustness section of the paper shows that this has no effect on the price elasticity estimates, ruling out this potential empirical problem (see robustness check number 8). Interestingly, Medicare expenditures per capita is not significantly related to utilization on the extensive margin, but it is positively and significantly related to utilization on the intensive margin. This is consistent with the hypothesis that the intensive margin reflects more of the physician's influence than the extensive margin.

Another important robustness check is to determine whether results are sensitive to how the out-of-pocket price is defined. Recall that the out-of-pocket price measure applied in this paper is calculated as the out-of-pocket expenditures for a family divided by the utilization amount for the family. This approach is useful for several reasons already mentioned. However, researchers have noted the challenges of selecting a proper measure for the out-of-pocket price (see Aron-Dine et al. (2013)), so it may be important to check the robustness of these findings to alternative definitions. One alternative measure looks at only an individual's out-of-pocket price, $\ln(OOPP_i)$, rather than the entire family's out-of-pocket price, $\ln(OOPP_f)$.[64] As another alternative, an out-of-pocket price variable is constructed using two years of claims data for the same family, which may be a better proxy for the marginal price in the long run.[65] For both robustness checks, the results remain unchanged.

Another issue is that the nonlinearity of the health insurance contracts creates a great deal of noise in the out-of-pocket price measure, which weakens the correlation between the instruments and the out-of-

[63] Another issues is that there may be moral hazard for individual behavior. For instance, those with more complete health insurance coverage may drive more recklessly or participate in more dangerous activity.

[64] When an individual's expenditures are zero the family's out-of-pocket price is used as the individual price. Results are reported in Table A10 of the appendix.

[65] This exercise implictly assumes that individuals change plans infrequently. Another reason to conduct this exercise is that several enrollees are dropped due to a lack of data on out-of-pocket expenditures. Using two years of data allows for an out-of-pocket price variable to be calculated for more individuals in the data. This result is reported in Table A10 of the appendix.

pocket price. As an alternative, one could look directly at the relationship between the MSA service price index and utilization to measure the full response to different service price indexes. One may view the full response involving the choices of the insurer, consumer, and employer. Although this analysis is complicated by the interpretation of the price coefficient,[66] a key advantage of this approach is that it removes the noise created by the nonlinear insurance contract, which arguably strengthens the identification of the price response. This analysis is conducted in the following subsection.

Additional robustness checks and a discussion of empirical issues are relegated to the appendix, including a discussion of how the empirical model is selected.

5.4.1 Empirical Relationship Between the Service Price Index and Utilization

Table 7 reports the effect of the MSA service price index, $\log(SPI^r)$, on utilization. Model 1 looks at this relationship directly, without any instruments, and shows an elasticity of -0.47. According to the first-stage estimates, approximately double the service price estimate is passed onto the consumer, which appears to lead to a near doubling of the price elasticities in Table 7 relative to the elasticities reported in Table 3. Although a selection problem does not arise in this analysis, the relationship between the unobserved quality of providers in the area and the service price index is a potential problem. The estimates for Models 2 through 5 use the familiar set of instruments. The results change slightly across IV strategies, with the elasticity for the preferred IV strategy in Model 4 increasing to -0.58.[67]

[66] i.e., how are the insurer, employer and the consumer jointly responding?

[67] One may be concerned with potential measurement error, since the aggregate price may not necessarily be relevant to an individual. As an alternative, an analysis using the estimated price paid by the family is calculated by dividing total disease expenditures by utilization, $SP_f = \frac{\sum_{i \in f} \sum_d c_{d,i}}{\sum_{i \in f} \sum_d SU_{d,i}}$. Qualitatively similar results are obtained when instruments are applied to this price measure.

Table 7. Effects of Log(SPI^r) on Overall Utilization (SU_i)

	(1)	(2)	(3)	(4)	(5)
Log(SPI^r)	-0.473***	-0.556***	-0.563***	-0.579***	-0.447***
	(-8.42)	(-4.31)	(-4.61)	(-5.46)	(-4.94)
Log(Median Income)	0.165***	0.158***	0.159***	0.158***	0.168***
	(4.29)	(3.44)	(3.06)	(3.18)	(3.74)
Log(Frac. Obese)	-0.00817	-0.0105	-0.00980	-0.0104	-0.00749
	(-0.47)	(-0.46)	(-0.47)	(-0.51)	(-0.32)
Log(Frac. Smokers)	0.0238	0.0220	0.0254	0.0251	0.0243
	(1.58)	(1.31)	(1.49)	(1.53)	(1.40)
Log(Frac. w/ College)	0.0164	0.0131	-0.00347	-0.00435	0.0174
	(0.63)	(0.43)	(-0.15)	(-0.19)	(0.55)
Residual Inclusion		0.245	0.139	0.308*	-0.0891
		(1.16)	(0.74)	(1.66)	(-0.46)
Number of Observations	8979207	8979207	8079984	8079984	8979207
Instruments	None	MSA Service Price Other Plans	Service Price of Other MSAs in State	Service Price of Other Plans & Other MSA Price	MSA Service Price, 25th Percentile

Notes: The z-statistics are in parentheses and are clustered by MSA. The z-statistics are computed using a bootstrap estimation that accounts for the two-stage estimation strategy. One, two, and three asterisks indicate significance at the 10-percent, 5-percent, or 1-percent significance level, respectively.

Since a large determinant of the amount an employer pays for medical care insurance will be determined by the medical service prices, one interpretation of Table 7 is that it proxies for the employer's elasticity response with respect to the price of insurance. With this interpretation, these estimates would suggest that employers are much more elastic than individuals, suggesting that much less generous plans are selected as service prices rise. Interestingly, these estimates are quite close to those of Gruber and Lettau (2004) that estimate an elasticity of insurance spending of -0.7 for firms.

Next, the analysis turns to the relationship between the service price and the weighted number of episodes shown in Table 8. Similar to the estimates in Table 7, the estimates show a statistically strong and negative relationship between the service price index and the weighted number of episodes across each of the alternative models. Again, the magnitude of the responsiveness approximately doubles relative to the corresponding estimates in Table 4.

Table 8. Effects of Log(SPI^r) on Weighted Number of Episodes ($Episodes^W_i$)

	(1)	(2)	(3)	(4)	(5)
Log(SPI^r)	-0.494***	-0.554***	-0.624***	-0.559***	-0.469***
	(-5.49)	(-5.04)	(-2.82)	(-3.73)	(-4.64)
Log(Median Income)	0.111***	0.106***	0.101***	0.107***	0.113***
	(4.39)	(4.29)	(2.64)	(3.18)	(4.35)
Log(Frac. Obese)	0.00686	0.00510	0.00957	0.0117	0.00753
	(0.39)	(0.27)	(0.49)	(0.58)	(0.37)
Log(Frac. Smokers)	-0.00738	-0.00872	-0.0108	-0.00928	-0.00688
	(-0.58)	(-0.57)	(-0.70)	(-0.64)	(-0.45)
Log(Frac. w/ College)	-0.0458**	-0.0481*	-0.0614**	-0.0591**	-0.0447*
	(-2.10)	(-1.95)	(-2.59)	(-2.55)	(-1.74)
Residual Inclusion		0.177	0.252	0.261	-0.0837
		(0.89)	(0.97)	(1.30)	(-0.53)
Number of Observations	8979207	8979207	8079984	8079984	8979207
Instruments	None	MSA Service Price Other Plans	Service Price of Other MSAs in State	Service Price of Other Plans & Other MSA Price	MSA Service Price, 25th Percentile

Notes: The z-statistics are in parentheses and are clustered by MSA. The z-statistics are computed using a bootstrap estimation that accounts for the two-stage estimation strategy. One, two, and three asterisks indicate significance at the 10-percent, 5-percent, or 1-percent significance level, respectively.

Table 9 reports the relationship between the disease-specific service price index and utilization per episode. Model 1 does not instrument for the service price and shows a negative and significant relationship between the service price index and the amount of utilization per episode. Model 2 of Table 9 includes an additional IV strategy that is instrumented using prices on other diseases, excluding disease d.[68] Models 3 through 6 contain the familiar set of instruments. Across all of the estimates, the price response along the intensive margin accounts for a relatively small fraction of the total price response. In all cases, the elasticity in Table 9 accounts for less than one fifth of the total price response reported in Table 7.

[68] Two price indexes are used as instruments: (1) an index built from other diseases in the same Major Practice Category (MPC) class; and (2) an index of diseases outside the same MPC class. For hypertension, this would mean that one price index would be constructed using prices from all other cardiovascular diseases, excluding hypertension. A second index would be constructed using all MPC categories, excluding cardiology conditions.

Table 9. Effects of Log(SPI^r_d) on Utilization per Episode ($SU_{d,i}$)

	(1)	(2)	(3)	(4)	(5)	(6)
Log(SPI^r_d)	-0.0649***	-0.0833*	-0.0857***	-0.103***	-0.102***	-0.0879**
	(-4.37)	(-1.94)	(-3.78)	(-2.63)	(-3.67)	(-2.22)
Log(Median Income)	0.00149	0.00290	-0.00727	-0.00191	-0.00874	0.00253
	(0.09)	(0.16)	(-0.39)	(-0.10)	(-0.46)	(0.14)
Log(Frac. Obese)	-0.0346***	-0.0341***	-0.0375***	-0.0407***	-0.0431***	-0.0342***
	(-4.21)	(-4.16)	(-4.20)	(-4.89)	(-4.71)	(-4.10)
Log(Frac. Smokers)	0.0192***	0.0191***	0.0178***	0.0212***	0.0200***	0.0190***
	(3.10)	(3.02)	(2.80)	(3.24)	(2.97)	(2.97)
Log(Frac. w/ College)	0.0311***	0.0296***	0.0266***	0.0281***	0.0244**	0.0295***
	(3.60)	(3.38)	(2.92)	(3.09)	(2.53)	(3.29)
Residual Inclusion		0.0218	-0.00816	0.0396	0.00536	0.0276
		(0.47)	(-0.33)	(1.06)	(0.18)	(0.66)
Number of Observations	28533318	27812331	23813449	25835741	21561379	27812331
Instruments	None	Disease-specific MSA Service Prices (Other Diseases)	Disease-Specific MSA Service Prices Other Plans	Disease-specific Service Prices of Other MSAs in State	Disease-specific,Service Price of Other Plans & Other MSA Prices	Disease-Specific MSA Service Price, 25th Percentile

Notes: The z-statistics are in parentheses and are clustered by MSA-MPC disease category. Due to the larger number of observations, the z-statistics are not adjusted for the two-stage estimation. However, applying a boostrap estimate to Model 5 that accounts for the two-stage estimation produces z-stats very similar to those reported in Model 5. One, two, and three asterisks indicate significance at the 10-percent, 5-percent, or 1-percent significance level, respectively.

Tables 7, 8 and 9 present estimates of the relationship between the service price index and the utilization measures. The linearity of both the service price index and the instruments greatly improves identification and allows for additional covariates to be incorporated into Tables 7, 8 and 9 without significantly weakening the instruments. Tables A7, A8, and A9 in the appendix repeat the analysis of Tables 7, 8, and 9, but include state fixed effects.[69] The overall elasticity estimates reported in Table A7 are similar to those in Table 7, although a bit more inelastic and less statistically significant. The results in Table 8 and 9 also change, with more of the price responsiveness shifting to the intensive margin and away from the extensive margin. The key take away from the estimates in Tables A7, A8, and A9 is that they demonstrate that price responsiveness may be identified using only within-state variation in service prices. However, the results should be interpreted with some caution, since state fixed effects are likely to remove important variation in the service price index across MSAs.

It is also worth highlighting that identification may be strengthened even further when examining utilization per episode. In particular, disease-specific service price indexes vary for each disease in each MSA, as documented by Dunn, Shapiro and Liebman (2012). Therefore, MSA fixed effects may be included and still identify price effects by using differences in disease-specific prices across MSAs. Results using MSA fixed effects are qualitatively similar to those reported in A9 of the appendix. This finding is important, since it highlights that identification may be achieved, even after removing all MSA-specific demand factors.

[69] The IV strategy using other prices in the state cannot be applied in this case.

6 Conclusion

This paper focuses on a fundamental empirical problem in the health literature: measuring consumer responsiveness to out-of-pocket price. To overcome the selection problems common in these studies, a unique approach is taken that exploits the large variation in negotiated prices of medical care services across areas. A service price index is used as an instruments that affects the medical costs of insurers and ultimately influences the out-of-pocket prices paid by consumers, but is not directly related to insurance selection. Applying this strategy, the demand estimates reveal that the consumer's response to out-of-pocket price is negative, significant and inelastic, with the main results mimicking those found in the RAND health insurance experiment. That is, after more than 30 years, the key results of the RAND study are reflected in observed variations in out-of-pocket price and utilization outside of the experimental setting. Moreover, the movements in negotiated service prices are shown to be closely correlated with out-of-pocket prices, demonstrating a clear mechanism for how changes in negotiated prices ultimately affect consumers and medical care utilization.

The instruments used in this study are not random and are determined by market forces. Consequently, the identification rests on the assumption that the underlying service price instruments are not related to the unobserved demand for insurance. Therefore, as in any econometric study that does not involve a random experiment, researchers cannot be absolutely certain that there is not an unobserved variable interfering with identification. However, in an expermental setting, researchers cannot be certain that their experimental design matches how consumers and markets behave outside of the experiment. Therefore, the main contribution of this paper should be viewed as complementary to experimental evidence, by revealing that the patterns observed in real world health care markets function how we might expect based on experimental evidence.

While demand elasticities reported in this paper are important, perhaps the greater contribution is providing a methodology for identifying medical care demand by connecting observed price movements to changes in utilization. This approach may be used to address questions not previously analyzed. One extension is to look at different components of medical care demand. For instance, one could look at the demand for physician services or drug services or substitution across these service categories. As a more challenging task, economists may be interested in the effect of new technologies on demand. That is, rather than focus on a cross-section, where similar medical technologies are available across markets, future work may benefit from employing panel data to study the impact of new technologies on consumer welfare.[70] As with nearly all studies of medical care demand, one limitation of this paper is that it ignores the potential price sensitivity of physicians. A key area for future research is to better understand the joint role of the consumer and physician price sensitivity and their effect on utilization, especially when analyzing utilization along the intensive margin (see Dunn and Shapiro (2012)). In all cases, the approach for studying demand applied here may be applied to other topics, which may lead to a deeper understanding of medical care markets.

[70] Rather than attempting to identify the value of new technologies in a time series (e.g., Dunn (2012) and Lucarelli and Nicholson (2009)), one could employ panel data, which may show different prices and rates of adoption across markets.

References

[1] Acemoglu, Daron, Amy Finkelstein, and Matthew Notowidigdo, (2013), “Income and Health Spending: Income from Oil Price Shocks”, *Review of Economics and Statistics*, Forethcoming.

[2] Aizcorbe, Ana and Nicole Nestoriak, (2011), “Changing Mix of Medical Care Services: Stylized Facts and Implications for Price Indexes”, *Journal of Health Economics*, 30 (3) pgs 568-574.

[3] Aron-Dine, Aviva, Liran Einav, Amy Finkelstein, and Mark Cullen, (2012), “Moral Hazard in Health Insurance: How Important is Forward Looking Behavior?”, NBER Working Paper No. 17802.

[4] Aron-Dine, Aviva, Liran Einav, and Amy Finkelstein, (2013), “The RAND Health Insurance Experiment, Three Decades Later”, *Journal of Economic Perspectives*, 27(1) pgs 1-28.

[5] Baicker, Katherine, Sarah Taubman, Heidi Allen, Mira Bernstein, Jonathan Gruber, Joseph Newhouse, Eric Schneider, Bill Wright, Alan Zaslavsky, Amy Finkelstein, (2013), “The Oregon Experiment - Effects of Medicaid on Clinical Outcomes”, *New England Journal fo Medicine*, 368 pgs 1713-1722.

[6] Buntin, Melinda Beeuwkes and Alan M. Zaslavsky, (2004), “Too Much Ado about Two-part Models and Transformation? Comparing Methods of Modeling Medicare Expenditures”, *Journal of Health Economics*, 23 pgs 525-542.

[7] Drury, CA and M. Louis, (2002), “Exploring the Association between Body Weight, Stigma of Obesity, and Health Care Avoidance”, *Journal of American Academy of Nurse Practitioners*, 14(12) pgs 554-61.

[8] Duarte, Fabian, (2012), “Price Elasticity of Expenditure Across Health Care Services”, *Journal of Health Economics*, 31 pgs 824-841.

[9] Dunn, Abe, Eli Liebman, and Adam Shapiro, (2012), “Decomposing Medical-Care Expenditure Growth”, BEA Working Paper (WP2012-10).

[10] Dunn, Abe, Adam Shapiro, and Eli Liebman, (2012), “Geographic Variation in Commercial Medical Care Expenditures: A Decomposition Between Price and Utilization”, BEA Working Paper.

[11] Dunn, Abe, Adam Shapiro, and Eli Liebman, (2012), “Technical Appendix: Geographic Variation in Commercial Medical Care Expenditures: A Decomposition Between Price and Utilization”, BEA Research Website: http://www.bea.gov/national/health_care_satellite_account.htm.

[12] Dunn, Abe and Adam Shapiro, (2012), “Physician Market Power and Medical-Care Expenditures”, BEA Working Paper.

[13] Finkelstein, Amy, (2007), “The Aggregate Effects of Health Insurance: Evidence from the Introduction of Medicare”, *Quarterly Journal of Economics*, 122(3) pgs 1-37.

[14] Finkelstein, Amy, Sarah Taubman, Bill Wright, Mira Bernstein, Jonathan Gruber, Joseph Newhouse, Heidi Allen, Katherine Baicker, (2012), “The Oregon Health Insurance Experiment: Evidence From the First Year”, The Quarterly Journal of Economics, 127(3) pgs 1057-1106.

[15] Eichner, Mattew, (1998), "The Demand for Medical Care: What People Pay Does Matter", *American Economic Review Papers and Proceedings*, 88(2) pgs 117-121.

[16] Einav, Liran, Amy Finkelstein, Stephan Ryan, Paul Schrimpf, and Mark R. Cullen, (2013), "Selection on Moral Hazard in Health Insurance", *American Economic Review*, 103(1) pgs 1-44.

[17] Gaynor, Martin and William Vogt, (2003), "Competition Among Hospitals", *RAND Journal of Economics*, 34(4) pgs 764-785.

[18] Goldman, D. and J. Smith, (2002), "Can Patient Self-Management Help Explain the SES Health Gradient?", *Proceedings of the National Academy of Sciences*, 99(16) pgs 10929-10934.

[19] Gottlieb, Daniel, Weiping Zhou, Yunjie Song, Kathryn Gilman Andrews, Jonathan Skinner and Jason Sutherland, (2010), "Prices Don't Drive Regional Medicare Spending Variations", *Health Affairs*, 29(3) pgs 537-543.

[20] Gruber, Jonathan and Michael Lettau, (2004), "How Elastic is the Firm's Demand for Health Insurance?", *Journal of Public Economics*, 88 pgs 1273-1293.

[21] Hausman, Jerry, (1996), "Valuation of New Goods Under Perfect and Imperfect Competition", in T. Bresnahan and R. Gordon, eds. The Economics of New Goods, Studies in Income and Wealth Vol 58, Chicago; National Bureau of Economic Research.

[22] Keeler, Emmett and John Rolph, (1988), "The Demand for Episodes of Treatment in Health Insurance Experiment", *Journal of Health Economics*, 7, pgs 337-367.

[23] Kowalski, Amanda, (2010), "Censored Quantile Instrumental Variable Estimates of the Price Elasticity of Expenditure on Medical Care", Working Paper.

[24] Manning, Willard G., Joseph Newhouse, Naihua Duan, Emmett Keeler, and Arleen Leibowitz, (1987), "Health Insurance and the Demand for Medical Care: Evidence from a Randomized Experiment", *American Economic Review*, 77(3) pgs 251-277.

[25] Manning, Willard G. and John Mullahy, (2001), "Estimating Log Models: To Transform or Not to Transform?", *Journal of Health Economics*, 20 pgs 461-494.

[26] Manning, Willard, Anirban Basu, and John Mullahy, (2005), "Generalized Modeling Approaches to Risk Adjustment of Skewed Outcomes Data", *Journal of Health Economics*, 24 pgs 465-488.

[27] Mullahy, John, (1998), "Much ado about two: reconsidering retransformation and the two-part model in health econometrics", *Journal of Health Economics*, 17 pgs 247-281.

[28] Nevo, Aviv, (2000), "A Practitioner's Guide to Estimation of Random-Coefficients Logit Models of Demand", *Journal of Economic and Management Strategy*, 9(4) pgs 513-548.

[29] Nevo, Aviv, (2001), "Measuring Market Power in the Ready-to-eat Cereal Industry", *Econometrica*, 69(2) pgs 307-342.

[30] Newhouse, Joseph, (1992), "Medical Care Costs: How Much Welfare Loss?", *Journal of Economic Perspectives*, 6(3) pgs 3-21.

[31] Newhouse, Joseph P., and the Insurance Experiment Group, (1993), *Free for All?* Cambridge: Harvard University Press.

[32] Newhouse, Joseph and Charles Phelps, (1976), "New Estimates of Price and Income Elasticities of Medical Care Services", The Role of Health Insurance in the Health Services Sector, Richard Rosett (ed.) (New York: Neal Watson).

[33] Sorensen, Alan, (2003), "Insurer-Hospital Bargaining: Negotiated Discounts in Post-deregulation Connecticut", *Journal of Industrial Economics*, 51(4) pgs 469-490.

[34] Skinner, Jonathan, (2012), "Causes and Consequences of Regional Variations in Health Care", *Handbook of Health Economics*, Chapter 2, pgs 45-93.

[35] Sundmacher, L, (2012), "The Effect of Health Shocks on Smoking and Obesity", *Europeon Journal of Health Economics*, 13(4) pgs 451-465.

[36] Town, Robert, and Gregory Vistnes, (2001), "Hospital Competition in HMO Networks", *Journal of Health Economics*, 20 pgs 733-753.

[37] Terza, Joseph, Anirban Basu, and Paul Rathouz, (2008), "Two-stage Residual Inclusion Estimation: Addressing Endogeneity in Health Econometric Modeling", *Journal of Health Economics*, 27 pgs 531-543.

7 Appendix - For Online Publication

7.1 Functional Form and Modeling Assumptions

The utilization data are highly skewed for all three measures of utilization. Applying a box-cox model to test for the appropriate functional form suggests a log transformation of the data, which greatly reduces the skewness.[71]

Applying a least squares model may be biased in the face of heteroskedasticity, so a Park test is applied to check for the presence of heteroskedasticity. The test is applied to each of the utilization measures, which shows a clear and strong relationship between the square of the least squares residuals and several of the independent variables. This finding suggests that heteroskedasticity is present and complex, favoring the application of GLM models. Next, tests are conducted to select the most appropriate GLM estimator. To assist in making this selection Manning and Mullahy (2001) suggest using a Park test to estimate the relationship between the mean of the predicted value and the variance of the error term. For all three components of utilization, the model suggests that the standard deviation is approximately proportional to the mean, implying a Gamma distribution, although the tests cannot reject the variance being proportional to the mean (i.e., Poisson distribution).

The Park tests suggest that GLM with a Gamma distribution may be preferred, but additional tests are conducted to determine how well the GLM Gamma and Poisson models fit the data (this discussion follows ideas from Buntin and Zaslavsky (2004)). As a first step, the two models are estimated on a 20 percent random sample, with the first model assuming a Poisson distribution and the second assuming

[71] For all three utilization equations, the box cox test finds the maximum-likelihood value of λ for the dependent variable: $ISU^{(\lambda)} = \frac{y^{\lambda}-1}{\lambda}$. The analysis shows an estimated value of λ near 0, indicating a log transformation.

a Gamma distribution. Each model's predicted value of utilization is computed for the remaining 80 percent of the data. Using these predicted values, the mean absolute prediction error and the mean square forecast errors are computed to determine the predictive accuracy of each model. The analysis shows that the fit of the GLM-Poisson model is considerably worse.[72] Given the size of the data, this test was not repeated hundreds of times by resampling, as in Buntin and Zaslavsky (2004). However, additional random samples were selected and analyzed and results remained unchanged. Although the Gamma distribution is preferred in the analysis, it is worth noting that the elasticity estimated on the extensive margin remains unchanged using either distributional assumption.

Another modeling decision was whether to apply a two-part model, which models two distinct decisions: (1) the decision to use any medical services; and (2) the amount of utilization to use conditional on utilizing some services. An investigation of the key estimates produced by the two-part model show that they are both quantitatively and qualitatively very similar to those produced by the GLM model using a Gamma distribution. Ultimately, the GLM model with the Gamma distribution is presented since the coefficients of the model are easier to interpret and the results are essentially unchanged.

7.2 Additional Robustness Checks

Below is a numerical list of additional robustness checks:

1. The functional form may be a concern for some readers. As a check, the elasticities are estimated using a two-part model. The two-part model consists of (1) a Probit model indicating whether utilization is positive; and (2) for positive utilization, a GLM model with a log link and Gamma distribution.(Table A10, Model 1).

2. Estimate the elasticity using individual out-of-pocket price ($OOPP_i$), rather than the family out-of-pocket price ($OOPP_f$) (Table A10, Model 2). In cases where $OOPP_i$ is not observed, the value $OOPP_f$ is used.

3. Estimate family out-of-pocket price using two years of claims data (Table A10, Model 3). Those individuals that have zero expenditures in both years are dropped from the analysis.

4. One concern with identifying price elasticities in a cross section is that a particular outlier MSA may greatly influence the elasticity estimates. To check if this is a concern, the sample is split, approximately in half. First, the sample is split by the number of enrollees in the MSA and results are qualitatively similar in each subsample (Table A10, Models 4 and 5). Next, the sample is split by region and, again, results are qualitatively similar in each subsample (Table A10, Models 6 and 7). The one anomaly is the low elasticity on the extensive margin when looking at the South and West region in Model 7. Upon further investigation, it appears that the low elasticity is caused by the inclusion of regional fixed effects that removes much of the variation necessary to identify an elasticity in this smaller subsample. When the regional fixed effects are removed, the elasticity estimates fall in the expected range and are significant in this subsample.

[72] These errors are computed using level predictions of utilization. If fit is measured using log utilization, the two models produce very similar fits.

5. The inclusion of regional dummies controls for region-specific utilization differences, but also removes across-region variation in service prices from the analysis. As another check, region fixed effects are removed from the specification and the main results are unchanged (Table A10, Model 8).

6. The MSA service price instrument, SPI^r, is calculated the same for all individuals in the data. However, the age and sex of individuals may make the expected disease treatment individual-specific. For instance, individuals in their 20s are less likely to have expenditures on heart-related conditions. An individual-specific MSA service price index is calculated for each individual, where the disease-specific service prices are weighted by the expenditure share of each disease for each individual's age, sex and family size category. The results are qualitatively similar to the other estimates (Table A10, Model 9).

7. For approximately eight percent of the individuals where expenditure information is not observed, the out-of-pocket price is imputed using expenditures from other individuals in the market. The individual categories used for the imputation include age, sex, plan-type, size of family, and data contributor for each MSA. The estimates that remove these imputations are qualitatively unchanged (Table 10, Model 10). Also note, that the estimates using two years of claims data to compute $OOPP_f$, (Table 10, Model 3), do not apply any imputation and results are also similar to the other estimates.

8. One may be concerned that the observed prices may be related to how physicians practice medicine differently across markets. To control for this possibility, a robustness check is conducted that accounts for the propensity of physicians to utilize medical services by including the log of Medicare expenditures per capita. Medicare expenditures per capita is not significantly related to utilization on the extensive margin (Table A10, Model 11), but it is positively and significantly related to utilization on the intensive margin (Table A11, Model 7). In all cases, the inclusion of the Medicare expenditures per capita has little effect on the elasticity estimates.

9. Around 13.8 percent of expenditures in the claims data are ungrouped and excluded from much of the analysis. As a check on whether dropping these expenditures has an effect, an alternative methodology for calculating utilization is applied that includes ungrouped expenditures. Specifically, overall utilization is calculated by dividing total expenditures by the individual-specific price index, SP_i, (i.e., $Adj.SU_i = \frac{\left(\sum_{d \in i} c_d\right) + \text{ungrouped expenditures}}{SP_i}$) to obtain a measure of utilization that includes the ungrouped claims. Using this alternative measure of utilization, similar elasticity estimates are obtained (Table A11, Model 1).

10. Apply a simple episode count ($Episodes_i$) rather than the weighted episode count ($Episodes_i^w$) (Table A11, Model 2).

11. Tables 7 and 8 look at the direct relationship between price indexes and utilization. One possible concern with looking at an overall price index is that it may capture the availability or adoption of different technologies in different areas. Given the variety of instruments applied this seems unlikely, but an additional robustness check is conducted using a "low-tech" service. Specifically, the average negotiated price for a 15-minute office visit to a general MD is used as an instrument for the MSA

service price index. The results of Tables 7 and 8 remain qualitatively unchanged (Table A11, Models 3 and 4).

12. To check for the importance of controlling for illness severity, controls are included to account for comorbidities and severity when analyzing utilization along the intensive margin, $SU_{d,i}$. The controls include dummy variables for the number of comorbidities (Table A11, Model 5).

13. MSA fixed effects are included in the analysis studying the effects of $SPI_{d,i}$ on $SU_{d,i}$ (Table A11, Model 6).

14. One might be concerned that there may be selection issues since individuals or firms may choose to be uninsured in markets with higher service prices. To check for this possibility, the county unemployment rate and the fraction of uninsured individuals were included in the analysis. There was no effect on the main results and each of these coefficients were insignificant. The effects were so small they are not included in the robustness tables. However, there is some evidence that higher prices are significantly related to a higher fraction of the population without insurance. Table A12 of the appendix reports a regression of $\log(SPI^r)$ on the fraction of individuals under 65 in the county without health insurance. The coefficient of the OLS regression is about 0.04, and the estimate is significant at the 90th percentile. The average rate of the uninsured population is about 0.17, which translates into an elasticity of around 0.24 (=0.04/0.17). This elasticity is very close to the insurance offer elasticity of Gruber and Lettau (2004) of 0.25. When instruments are applied to account for possible measurement error in the service price, the elasticity is considerably higher, 0.41 (0.07/0.17). This analysis is not conducted at the firm or employee level, so this analysis is only suggestive of general relationship and is not intended to be a precise estimate.

7.3 Reported Full Estimates

Table A1.1 Effects of Out-of-pocket-price on Utilization - Full Estimates

	Overall Utilization (SU_i)		Weighted Num. of Episodes ($Episodes^W_i$)			(Continued)			
	(1)	(2)	(3)	(4)		(1)	(2)	(3)	(4)
Log($OOPC_i$)	-0.223***	-0.199***	-0.228***	-0.197***	Frac. of Hosp. Med. Schools	-0.0228	-0.0125	-0.0107	-0.00124
	(-3.47)	(-3.75)	(-3.91)	(-3.19)		(-1.26)	(-0.64)	(-0.60)	(-0.07)
Log(Median Income)	0.189***	0.198***	0.118***	0.127***	Data Source: Insurer Data	0.0114	0.0124	0.0399***	0.0369**
	(3.57)	(4.97)	(3.70)	(3.31)		(0.58)	(0.63)	(2.59)	(2.51)
Log(Frac. Obese)	-0.0346	-0.0231	-0.0151	-0.00153	PPO	-0.140***	-0.154***	-0.125***	-0.136***
	(-1.50)	(-0.95)	(-0.65)	(-0.07)		(-3.85)	(-5.26)	(-3.79)	(-5.19)
Log(Frac. Smokers)	0.0214	0.0260	-0.00345	-0.00247	POS	-0.244***	-0.260***	-0.206***	-0.213***
	(1.35)	(1.53)	(-0.26)	(-0.20)		(-7.44)	(-8.87)	(-7.72)	(-9.18)
Log(Frac. w/ College)	0.0368	0.0287	-0.0306	-0.0364	Comprehensive	-0.0991***	-0.116***	-0.0849***	-0.0961***
	(1.27)	(1.13)	(-1.32)	(-1.51)		(-2.95)	(-4.25)	(-3.02)	(-3.80)
Residual Inclusion	-0.408***	-0.426***	-0.0693	-0.0986*	High Deductible Health Plan	-0.198***	-0.226***	-0.258***	-0.283***
	(-7.01)	(-9.34)	(-1.24)	(-1.71)		(-5.06)	(-7.36)	(-7.46)	(-8.11)
Log(Rent)	-0.0964	-0.0739	0.0467	0.0645	Family Size=2	0.0704***	0.0714***	0.0465***	0.0471***
	(-1.22)	(-1.19)	(0.75)	(1.06)		(8.01)	(8.56)	(6.01)	(6.01)
Male	0.103***	0.102***	0.0635***	0.0624***	Family Size=3	0.0270***	0.0297***	0.0145**	0.0157**
	(16.72)	(16.72)	(18.79)	(14.96)		(3.61)	(4.09)	(1.99)	(2.05)
Age 17 to 24	0.326***	0.320***	0.312***	0.303***	Family Size=4	-0.0680***	-0.0655***	-0.0553***	-0.0531***
	(35.36)	(30.19)	(28.11)	(29.42)		(-9.02)	(-7.75)	(-7.54)	(-5.82)
Age 25 to 34	0.795***	0.794***	0.766***	0.762***	Family Size>=5	-0.164***	-0.160***	-0.154***	-0.150***
	(74.30)	(65.08)	(64.92)	(69.91)		(-21.72)	(-20.03)	(-18.36)	(-18.20)
Age 35 to 44	0.832***	0.830***	0.752***	0.750***	Year	0.0261***	0.0233**	0.0217***	0.0194***
	(84.98)	(84.61)	(72.31)	(94.94)		(3.20)	(2.44)	(3.88)	(3.27)
Age 45 to 54	1.023***	1.021***	0.964***	0.965***	New England	0.0585*	0.0613	0.0302	0.0249
	(81.84)	(85.80)	(79.67)	(104.89)		(1.70)	(1.62)	(0.77)	(0.57)
Age 55 to 64	1.272***	1.270***	1.252***	1.251***	Mid-Atlantic	0.0501**	0.0494**	0.0616*	0.0575*
	(85.37)	(88.81)	(88.17)	(118.02)		(2.01)	(2.18)	(1.66)	(1.87)
Age 17 to 24 * Male	-0.546***	-0.539***	-0.602***	-0.593***	East North Central	0.0409	0.0466*	0.0492	0.0521
	(-33.09)	(-30.98)	(-43.00)	(-53.91)		(1.38)	(1.86)	(1.27)	(1.24)
Age 25 to 34 * Male	-0.942***	-0.943***	-0.897***	-0.895***	West North Central	0.0112	0.0429	0.0250	0.0377
	(-73.02)	(-73.67)	(-76.02)	(-77.83)		(0.33)	(1.14)	(0.67)	(1.06)
Age 35 to 44 * Male	-0.587***	-0.586***	-0.511***	-0.509***	South Atlantic	0.0783***	0.0775***	0.0955***	0.0914***
	(-51.49)	(-56.89)	(-75.26)	(-76.77)		(3.66)	(3.64)	(3.00)	(3.92)
Age 45 to 54 * Male	-0.356***	-0.354***	-0.311***	-0.310***	East South Central	0.117*	0.117	0.145***	0.133***
	(-38.28)	(-35.79)	(-43.86)	(-43.06)		(1.82)	(1.57)	(3.30)	(3.28)
Age 55 to 64 * Male	-0.174***	-0.174***	-0.150***	-0.150***	West South Central	0.0944***	0.0924**	0.0680**	0.0614**
	(-14.50)	(-16.89)	(-24.31)	(-22.97)		(4.20)	(3.34)	(2.49)	(2.41)
Number of Observations	8979207	8079984	8979207	8079984					

Table A1.2 Effects of Out-of-pocket-price on Utilization - Full Estimates

	Service Utilization Per Episode ($SU_{i,d}$)			(Continued)	
	(1)	(2)		(1)	(2)
Log($OOPC_f$)	-0.0252	-0.0503*	Frac. of Hosp. Med. Schools	-0.0175**	-0.0190*
	(-1.29)	(-1.80)		(-2.04)	(-1.79)
Log(Median Income)	0.00644	-0.00937	Data Source: Insurer Data	-0.0854***	-0.0772***
	(0.34)	(-0.48)		(-15.11)	(-12.17)
Log(Frac. Obese)	-0.0374***	-0.0492***	PPO	-0.0923***	-0.0895***
	(-4.15)	(-5.02)		(-9.45)	(-7.67)
Log(Frac. Smokers)	0.0184***	0.0187**	POS	-0.111***	-0.115***
	(2.67)	(2.56)		(-12.10)	(-11.27)
Log(Frac. w/ College)	0.0304***	0.0252**	Comprehensive	-0.0721***	-0.0746***
	(3.06)	(2.21)		(-6.77)	(-6.16)
Residual Inclusion	-0.177***	-0.125***	High Deductible Health Plan	-0.0457***	-0.0411**
	(-9.25)	(-4.59)		(-3.75)	(-2.55)
Log(Rent)	-0.0637**	-0.0386	Family Size=2	-0.00309	-0.00797***
	(-2.18)	(-1.20)		(-1.41)	(-2.73)
Male	0.0208***	0.0193***	Family Size=3	-0.0299***	-0.0335***
	(6.72)	(5.66)		(-11.77)	(-9.71)
Age 17 to 24	0.0611***	0.0677***	Family Size=4	-0.0571***	-0.0605***
	(10.18)	(9.52)		(-21.36)	(-17.27)
Age 25 to 34	0.0462***	0.0541***	Family Size>=5	-0.0636***	-0.0713***
	(4.95)	(4.84)		(-19.67)	(-15.89)
Age 35 to 44	0.140***	0.151***	Year	0.0159***	0.0184***
	(10.13)	(9.07)		(6.72)	(6.07)
Age 45 to 54	0.172***	0.187***	New England	0.0387**	0.0475**
	(9.31)	(8.29)		(1.97)	(1.98)
Age 55 to 64	0.175***	0.188***	Mid-Atlantic	0.0128	0.0220
	(7.73)	(6.85)		(0.74)	(1.18)
Age 17 to 24 * Male	-0.00520	-0.0116**	East North Central	0.0270	0.0338
	(-1.01)	(-2.09)		(1.57)	(1.61)
Age 25 to 34 * Male	-0.000374	0.0118*	West North Central	0.0160	0.0364*
	(-0.07)	(1.84)		(0.93)	(1.90)
Age 35 to 44 * Male	-0.0294***	-0.0175***	South Atlantic	0.00455	0.0133
	(-7.27)	(-3.81)		(0.36)	(0.94)
Age 45 to 54 * Male	-0.0179***	-0.0122**	East South Central	0.0114	0.0304*
	(-4.41)	(-2.50)		(0.69)	(1.66)
Age 55 to 64 * Male	0.00142	0.00458	West South Central	0.0454***	0.0563***
	(0.38)	(1.09)		(3.13)	(3.45)
Number of Observations	27812331	21561379			
Instruments	MSA Service Price	Service Price of Other Plans & Other MSA Price			

Notes: The z-statistics are in parentheses and are clustered by MSA-MPC disease category. Due to the larger number of observations, the z-statistics are not adjusted for the two-stage estimation. However, applying a boostrap estimate to Model 5 that accounts for the two-stage estimation produces z-stats very similar to those reported in Model 5. One, two, and three asterisks indicate significance at the 10-percent, 5-percent, or 1-percent significance level, respectively.

7.4 Disease-Specific Out-of-pocket Price Response

Table A2. Effect of Out-of-pocket Price on Utilization - Probit Model

	Diabetes	Hypertension	High Cholesterol	Depression	Migraine	Appendicitis
Log($OOPP_f$)	-0.0517**	-0.112**	-0.107*	-0.132***	-0.0887**	0.0211
	(-2.20)	(-2.43)	(-1.76)	(-2.95)	(-2.43)	(0.56)

Notes: The z-statistics are in parentheses and are clustered by MSA. One, two, and three asterisks indicate significance at the 10-percent, 5-percent, or 1-percent significance level, respectively.

7.5 Estimates with State Fixed Effects

Table A7. Relationship Between Log(SPI^r) and Overall Utilization (SU_i) with State FE

	(1)	(2)	(3)
Log(SPI^r)	-0.290***	-0.304**	-0.355***
	(-2.67)	(-2.29)	(-3.20)
Log(Median Income)	0.186***	0.185***	0.183***
	(4.61)	(5.04)	(5.20)
Log(Frac. Obese)	-0.0174	-0.0172	-0.0164
	(-1.05)	(-0.91)	(-0.92)
Log(Frac. Smokers)	0.0337***	0.0336**	0.0331**
	(2.65)	(2.02)	(2.04)
Log(Frac. w/ College)	0.00399	0.00397	0.00385
	(0.17)	(0.20)	(0.19)
Residual Inclusion		0.0201	0.210
		(0.10)	(1.22)
Number of Observations	8979207	8979207	8979207
Instruments	None	MSA Service Price Other Plans	MSA Service Price, 25th Percentile

Notes: The z-statistics are in parentheses and are clustered by MSA. The z-statistics are computed using a bootstrap estimation that accounts for the two-stage estimation strategy. One, two, and three asterisks indicate significance at the 10-percent, 5-percent, or 1-percent significance level, respectively.

Table A8. Relationship Between Log(SPI^r) and Weighted Number of Episodes ($Episodes^W_i$) with State FE

	(1)	(2)	(3)
Log(SPI^r)	-0.147*	-0.135	-0.120
	(-1.65)	(-1.39)	(-1.28)
Log(Median Income)	0.166***	0.167***	0.167***
	(5.37)	(5.86)	(5.80)
Log(Frac. Obese)	-0.00426	-0.00444	-0.00464
	(-0.29)	(-0.39)	(-0.43)
Log(Frac. Smokers)	0.0191**	0.0192**	0.0193**
	(2.11)	(1.99)	(2.11)
Log(Frac. w/ College)	-0.0137	-0.0137	-0.0136
	(-0.77)	(-1.20)	(-1.17)
Residual Inclusion		-0.0177	-0.0867
		(-0.14)	(-0.61)
Number of Observations	8979207	8979207	8979207
Instruments	None	MSA Service Price Other Plans	MSA Service Price, 25th Percentile

Notes: The z-statistics are in parentheses and are clustered by MSA. The z-statistics are computed using a bootstrap estimation that accounts for the two-stage estimation strategy. One, two, and three asterisks indicate significance at the 10-percent, 5-percent, or 1-percent significance level, respectively.

Table A9. Relationship between Log(SPI^r) and Utilization per Episode ($SU_{d,i}$) with State FE

	(1)	(2)	(3)	(4)	(5)	(6)
Log(SPI^r)	-0.0802***	-0.162***	-0.108***	-0.182***	-0.132***	-0.242***
	(-4.88)	(-2.92)	(-4.06)	(-3.67)	(-4.21)	(-4.08)
Log(Median Income)	-0.00468	-0.00536	-0.0115	-0.00631	-0.0112	-0.00749
	(-0.28)	(-0.31)	(-0.63)	(-0.36)	(-0.60)	(-0.44)
Log(Frac. Obese)	-0.0304***	-0.0291***	-0.0324***	-0.0313***	-0.0339***	-0.0284***
	(-4.31)	(-4.09)	(-4.25)	(-4.29)	(-4.30)	(-4.02)
Log(Frac. Smokers)	0.00584	0.00600	0.00571	0.00624	0.00621	0.00579
	(1.17)	(1.17)	(1.08)	(1.19)	(1.11)	(1.14)
Log(Frac. w/ College)	0.0149*	0.0132	0.0136	0.0164*	0.0150	0.0135
	(1.75)	(1.52)	(1.51)	(1.84)	(1.60)	(1.56)
Residual Inclusion		0.0880	-0.0125	0.110**	0.0176	0.176***
		(1.48)	(-0.56)	(2.44)	(0.62)	(2.94)
Number of Observations	28533318	27812331	23813449	25835741	21561379	27812331
Instruments	None	Disease-specific MSA Service Prices (Other Diseases)	Disease-Specific MSA Service Prices Other Plans	Disease-specific Service Prices of Other MSAs in State	Disease-specific, Service Price of Other Plans & Other MSA Prices	Disease-Specific MSA Service Price, 25th Percentile

Notes: The z-statistics are in parentheses and are clustered by MSA-MPC disease category. The z-statistics are computed using a bootstrap estimation that accounts for the two-stage estimation strategy. One, two, and three asterisks indicate significance at the 10-percent, 5-percent, or 1-percent significance level, respectively.

Table A10. Estimated Robustness Checks on SU_i and $Episode^w_i$

	Price Elasticity Estimate on SU_i	Price Elasticity Estimate on $Episode^w_i$
1. Two-Part Model: Effect of $Log(OOPP_f)$		
Probit	-0.141***	-0.141***
	(-2.87)	(-2.87)
GLM (log link & gamm distribution)	-0.156**	-0.151***
	(-2.25)	(-3.09)
Combined Effect	-0.198***	-0.192***
	(-2.79)	(-3.78)
2. Individual Out-of-pocket Price: Effect of $Log(OOPP_i)$	-0.158**	-0.194***
	(-2.18)	(-3.29)
3. Family Out-of-pocket Price Two Years: Effect of $Log(OOPP_i)$	-0.249***	-0.191***
	(-3.73)	(-4.07)
4. MSAs Large Enrollment (over 40,000): Effect of $Log(OOPP_f)$	-0.176**	-0.227***
	(-2.34)	(-3.49)
5. MSAs Small Enrollment (under 40,000): Effect of $Log(OOPP_f)$	-0.241*	-0.181**
	(-1.85)	(-2.38)
6. Regions NE and MW: Effect of $Log(OOPP_f)$	-0.213***	-0.257***
	(-2.62)	(-4.05)
7. Regions S and W: Effect of $Log(OOPP_f)$	-0.223*	-0.0903
	(-1.75)	(-1.19)
8. Exclude Regional Fixed Effects: Effect of $Log(OOPP_f)$	-0.251***	-0.270***
	(-3.00)	(-5.63)
9. Individual-Specific Expected SPI^r: Effect of $Log(OOPP_f)$	-0.280***	-0.275***
	(-3.67)	(-6.83)
10. No $OOPP_f$ imputation: Effect of $Log(OOPP_f)$	-0.179**	-0.177***
	(-2.52)	(-3.30)
11. Include log(Medicare Exp. Per Capita): Effect of $Log(OOPP_f)$	-0.182***	-0.186**
	(-2.74)	(-3.28)
Coefficient on Log(Medicare Exp. Per Capita)	0.127*	0.0761
	(1.72)	(1.26)

Notes: The z-statistics are in parentheses and are clustered by MSA. One, two, and three asterisks indicate significance at the 5-percent, 1-percent, or 0.1-percent significance level, respectively. Unless specified otherwise, the IV strategy will include two instruments: (1) a price index constructed from prices of other plans in the MSA and (2) a price index constructed from prices in other MSAs in the state. These estimates are based on a 30 percent random sample of the data. The estimated z-statistics are not adjusted for the first stage estimates of the residual inclusion. Accounting for the first stage estimates when calculating standard errors had only very small effects on the z-statistic estimates.

Table A11. Additional Robustness Checks

	Price Elasticity Estimate
1. Include Ungrouped Expenditures: Effect of Log($OOPP_f$) on Adjusted SU_i	-0.280***
	(-2.79)
2. Effect of Log($OOPP_f$) on Simple Episode Count ($Episode_i$)	-0.141***
	(-3.00)
3. Average 15-minute Visit Price Instrument: Effect of Log(SPI') on SU_i	-0.554***
	(-3.43)
4. Average 15-minute Visit Price Instrument: Effect of Log(SPI') on $Episode''_i$	-0.678***
	(-3.73)
5. Include Additional Severity Controls: Effect of Log($OOPP_f$) on $SU_{d,i}$	-0.0261
	(-0.89)
6. Include MSA Fixed-Effects: Effect of Log(SPI_d^r) on $SU_{d,i}$	-0.122***
	(-4.13)
7. Include log(Medicare Exp. Per Capita): Effect of Log($OOPP_f$) on $SU_{d,i}$	-0.0311
	(-1.06)
Coefficient on Log(Medicare Exp. Per Capita)	0.0864***
	(3.37)

Notes: The z-statistics are in parentheses and are clustered by MSA. One, two, and three asterisks indicate significance at the 10-percent, 5-percent, or 1-percent significance level, respectively. Unless specified otherwise, the IV strategy will include two instruments: (1) a price index constructed from prices of other plans in the MSA and (2) a price index constructed from prices in other MSAs in the state. These estimates are based on a 30 percent random sample of the data. The estimated z-statistics are not adjusted for the first stage estimates of the residual inclusion. Accounting for the first stage estimates when calculating standard errors had only very small effects on the z-statistic estimates.

Table A12. Effect of Log(SPI^r) on Fraction of County Population that is Uninsured

	(1)	(2)	(3)
Log(SPI^r)	0.0497*	0.0381*	0.0725***
	(1.95)	(1.91)	(2.92)
Log(Income)	-0.142***	-0.0741***	-0.0670***
	(-6.44)	(-4.23)	(-3.55)
log(Fraction w/ College)	-0.0404***	-0.0219***	-0.0230***
	(-5.57)	(-4.17)	(-4.00)
Log(Rent)	0.115***	0.0736***	0.0774***
	(5.42)	(4.09)	(4.00)
Log(Frac. Obesity)	-0.0340***	-0.0228***	-0.0205***
	(-4.82)	(-4.33)	(-3.61)
Log(Frac. Smokers)	-0.0162**	-0.00950*	-0.0119**
	(-3.07)	(-2.46)	(-2.90)
Share of Population over 65	-0.00856	0.477***	0.467**
	(-0.04)	(3.38)	(3.05)
Log(Median Age)	-0.0930	-0.194***	-0.183**
	(-1.43)	(-3.85)	(-3.25)
Year	0.00389	-0.000313	-0.00240
	(1.94)	(-0.20)	(-1.38)
Region Fixed Effects	No	Yes	Yes
IV Regression	No	No	Yes
N	740	740	627

Notes: The z-statistics are in parentheses and are clustered by county. One, two, and three asterisks indicate significance at the 10-percent, 5-percent, or 1-percent significance level, respectively. The dependant variable is from the ARF data and is the fraction of the county population under 65 that is uninsured. The instruments include the service price index from other MSAs in the state and a service price index constructed from the 25th percentile price information.